

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 49.4422 Seconds
(without alignments)
2917.574 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPICNLGHNPNVNSIPPAP.....SNSAVNTSNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organalle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2624.5	92.1	558	16 Q9R396	Q9R396 escherichia
2	2608.5	91.5	558	2 Q85506	Q85506 escherichia
3	1774.5	62.2	538	2 Q47014	Q47014 escherichia
4	1766.5	62.0	538	2 Q47016	Q47016 escherichia
5	1753.5	61.5	538	2 Q85508	Q85508 escherichia
6	1570	55.1	551	2 Q68258	Q68258 escherichia
7	1554.5	54.5	552	2 Q8KWH9	Q8KWH9 escherichia
8	1548.5	54.3	550	2 Q52147	Q52147 escherichia
9	1522	53.4	547	2 Q9WXK1	Q9WXK1 escherichia
10	1521	53.3	547	2 Q9ETI1	Q9ETI1 citrobacter
11	1495	52.4	549	2 Q50190	Q50190 escherichia
12	190	6.7	1323	5 Q9VI63	Q9VI63 drosophila
13	190	6.7	1376	5 Q9BZ29	Q9BZ29 drosophila
14	184	6.5	1321	5 Q962D1	Q962D1 drosophila
15	183	6.4	1015	5 Q8SZW9	Q8SZW9 drosophila
16	170	6.0	1254	5 Q94185	Q94185 caenorhabdi

17	165.5	5.8	2232	5 Q8IFX6	Q8IFX6 caenorhabdi
18	162.5	5.7	1291	5 Q77261	Q77261 drosophila
19	162.5	5.7	1300	5 Q9W5E0	Q9W5E0 drosophila
20	161	5.6	1353	5 Q9V8Q2	Q9V8Q2 drosophila
21	156	5.5	2921	5 Q9N973	Q9N973 leishmania
22	152.5	5.3	973	16 Q8XDQ4	Q8XDQ4 escherichia
23	152.5	5.3	1412	5 Q9VKJ1	Q9VKJ1 drosophila
24	149.5	5.2	676	5 Q9VB94	Q9VB94 drosophila
25	149.5	5.2	1241	5 Q8MNI0	Q8MNI0 dictyosteli
26	149	5.2	2276	2 Q93TY6	Q93TY6 staphylococ
27	149	5.2	2310	16 Q8CMU7	Q8CMU7 staphylococ
28	148.5	5.2	606	3 Q9P319	Q9P319 neurospora
29	148	5.2	1192	5 Q9TW45	Q9TW45 caenorhabdi
30	148	5.2	1192	5 Q17346	Q17346 caenorhabdi
31	147.5	5.2	3012	5 Q97205	Q97205 leishmania
32	145.5	5.1	1550	5 Q9W002	Q9W002 drosophila
33	145	5.1	3848	2 P94772	P94772 erwinia chr
34	145	5.1	4782	11 Q8KIG6	Q8KIG6 mus musculu
35	144.5	5.1	2585	5 Q23587	Q23587 caenorhabdi
36	144	5.1	1323	6 Q77509	Q77509 bos taurus
37	143.5	5.0	1275	5 Q76602	Q76602 caenorhabdi
38	143.5	5.0	1805	11 Q63661	Q63661 rattus norv
39	143	5.0	2678	5 Q9NDS4	Q9NDS4 dictyosteli
40	143	5.0	3443	11 Q8JZM8	Q8JZM8 mus musculu
41	142.5	5.0	1026	5 Q9V8B9	Q9V8B9 drosophila
42	142.5	5.0	3013	16 Q8ZA73	Q8ZA73 yersinia pe
43	142	5.0	2283	2 Q8VQ99	Q8VQ99 staphylococ
44	141.5	5.0	1027	5 Q9W3U2	Q9W3U2 drosophila
45	141.5	5.0	1096	5 Q17368	Q17368 caenorhabdi

ALIGNMENTS

RESULT 1

Q9R396 ID Q9R396 PRELIMINARY; PRT; 558 AA.
AC Q9R396;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-OCT-2002 (TREMBlrel. 22, Last annotation update)
DE Translocated intimin receptor Tir (Putative translocated intimin receptor protein).
GN TIR OR Z5112 OR ECS4561.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=86/24;
RX MEDLINE=99242825; PubMed=10225900;
RA Devlin R., Stein M., Reinscheid D., Abe A., Ruschowski S.,
RA Finlay B.B.;
RT "Enterohemorrhagic Escherichia coli O157:H7 produces Tir, which is translocated to the host cell membrane but is not tyrosine phosphorylated.";
RT Infect. Immun. 67:2389-2398(1999).
RL [2]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=EDL933;
RX MEDLINE=98339885; PubMed=9673266;
RA Perna N.T., Mayhew G.F., Posfai G., Elliott S., Donnenberg M.S.,
RA Kaper J.B., Blattner F.R.;
RT "Molecular evolution of a pathogenicity island from enterohemorrhagic Escherichia coli O157:H7.";
RT Infect. Immun. 66:3810-3817(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli O157:H7; STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

```

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RA "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
RT Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli O157:H7; STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki K., Ogasawara N., Yasunaga T.,
RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
RA "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22(2001).
DR EMBL; AF125993; AAD29391.1; -;
DR EMBL; AF071034; AAC31506.1; -;
DR EMBL; AE005595; AAG58825.1; -;
DR EMBL; AP002566; BAB37984.1; -;
DR InterPro; IPR003536; TIR receptor.
DR Pfam; PF03549; TIR receptor; 1.
DR PRINTS; PR01370; TRNSINTIMINR.
DR Receptor; Complete proteome.
KW Receptor; Complete proteome.
SQ SEQUENCE 558 AA; 58022 MW; 99C41722D2D4B4AA1 CRC64;

Query Match 92.1%; Score 2624.5; DB 16; Length 558;
Best Local Similarity 94.0%; Pred. No. 1e-135;
Matches 528; Conservative 3; Mismatches 24; Indels 7; Gaps 3;

Qy 1 MPIGNLGNPNVNSIPPAPLPSTQDAGRGQLINSTGPGSRALFTPVNSMADSGD 60
Db 1 MPIGNLGNPNVNSIPPAPLPSTQDAGRGQLINSTGPGSRALFTPVNSMADSGD 60
Qy 61 NRASDVPLGPNMRLAASEITLNDGFEVLHDHGLDNLNRQIGSSVFRVETQEDGKHA 120
Db 61 NRASDVPLGPNMRLAASEITLNDGFEVLHDHGLDNLNRQIGSSVFRVETQEDGKHA 120
Qy 121 VQGRNGVETSVVLSQDYARLQSIDPEGKDFVFTGGRGAGHAMVTVASDITEARQIL 180
Db 121 VQGRNGVETSVVLSQDYARLQSIDPEGKDFVFTGGRGAGHAMVTVASDITEARQIL 180
Qy 181 ELLEPKGTGESKAGESKVGELRESNCAENTTETQSTSSLSRSDPKLWALGTAT 240
Db 181 ELLEPKGTGESKAGESKVGELRESNCAENTTETQSTSSLSRSDPKLWALGTAT 240
Qy 241 GLIGLAATGIVQALALTPEDSPPTTDPDAASATETATRDQLTKEAFQNPQKNWIDE 300
Db 241 GLIGLAATGIVQALALTPEDSPPTTDPDAASATETATRDQLTKEAFQNPQKNWIDE 300
Qy 301 LGNAIPSGVLKDDVVANIEQAKAAGEAKQQAENNAQAOKKYDEQAKRQBELKVSSG 360
Db 301 LGNAIPSGVLKDDVVANIEQAKAAGEAKQQAENNAQAOKKYDEQAKRQBELKVSSG 360
Qy 361 AGYGLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTTSARTVENKPAANTPAQG 420
Db 361 AGYGLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTTSARTVENKPAANTPAQG 420
Qy 421 NVDTPGSEDTMESRRSSMASTSTFTDTSIG---GPCRIRMLMLKRCMTRCRLILLI 477
Db 421 NVDTPGSEDTMESRRSSMASTSTFTDTSIG---GPCRIRMLMLKRCMTRCRLILLI 477
Qy 478 RLFRIMGIQISVYVSTIQHPPRDTTNGARLLGNPSAGIQSTYARLALSGGLRHDWGGLT 537
Db 478 RLFRIMGIQISVYVSTIQHPPRDTTNGARLLGNPSAGIQSTYARLALSGGLRHDWGGLT 537
Qy 480 QNM---GNTDSVVYSTIQHPPRDTTNGARLLGNPSAGIQSTYARLALSGGLRHDWGGLT 536
Db 480 QNM---GNTDSVVYSTIQHPPRDTTNGARLLGNPSAGIQSTYARLALSGGLRHDWGGLT 536
Qy 538 GGSNSAVNTSNPPAPGSHRFV 559
Db 538 GGSNSAVNTSNPPAPGSHRFV 559
Qy 537 GGSNSAVNTSNPPAPGSHRFV 558
Db 537 GGSNSAVNTSNPPAPGSHRFV 558

```

```

RESULT 2
O85506 PRELIMINARY; PRT; 558 AA.
AC O85506;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Translocated intimin receptor TIR.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=95SP2;
RX MEDLINE=99003184; PubMed=9784578; Paton J.C.;
RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RA "Translocated intimin receptors (Tir) of shiga-toxicogenic Escherichia
RT coli isolates belonging to serogroups O26, O111, and O157 react with
RT sera from patients with hemolytic-uremic syndrome and exhibit marked
RT sequence heterogeneity.";
RL Infect. Immun. 66:5580-5586(1998).
DR EMBL; AF070067; AAC69314.1; -;
DR InterPro; IPR003536; TIR receptor.
DR Pfam; PF03549; TIR receptor; 1.
DR PRINTS; PR01370; TRNSINTIMINR.
KW Receptor.
SQ SEQUENCE 558 AA; 58175 MW; CA2CDDAC94527C2E CRC64;

Query Match 91.5%; Score 2608.5; DB 2; Length 558;
Best Local Similarity 93.4%; Pred. No. 7.6e-135;
Matches 525; Conservative 3; Mismatches 27; Indels 7; Gaps 3;

Qy 1 MPIGNLGNPNVNSIPPAPLPSTQDAGRGQLINSTGPGSRALFTPVNSMADSGD 60
Db 1 MPIGNLGNPNVNSIPPAPLPSTQDAGRGQLINSTGPGSRALFTPVNSMADSGD 60
Qy 61 NRASDVPLGPNMRLAASEITLNDGFEVLHDHGLDNLNRQIGSSVFRVETQEDGKHA 120
Db 61 NRASDVPLGPNMRLAASEITLNDGFEVLHDHGLDNLNRQIGSSVFRVETQEDGKHA 120
Qy 121 VQGRNGVETSVVLSQDYARLQSIDPEGKDFVFTGGRGAGHAMVTVASDITEARQIL 180
Db 121 VQGRNGVETSVVLSQDYARLQSIDPEGKDFVFTGGRGAGHAMVTVASDITEARQIL 180
Qy 181 ELLEPKGTGESKAGESKVGELRESNCAENTTETQSTSSLSRSDPKLWALGTAT 240
Db 181 ELLEPKGTGESKAGESKVGELRESNCAENTTETQSTSSLSRSDPKLWALGTAT 240
Qy 241 GLIGLAATGIVQALALTPEDSPPTTDPDAASATETATRDQLTKEAFQNPQKNWIDE 300
Db 241 GLIGLAATGIVQALALTPEDSPPTTDPDAASATETATRDQLTKEAFQNPQKNWIDE 300
Qy 301 LGNAIPSGVLKDDVVANIEQAKAAGEAKQQAENNAQAOKKYDEQAKRQBELKVSSG 360
Db 301 LGNAIPSGVLKDDVVANIEQAKAAGEAKQQAENNAQAOKKYDEQAKRQBELKVSSG 360
Qy 361 AGYGLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTTSARTVENKPAANTPAQG 420
Db 361 AGYGLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTTSARTVENKPAANTPAQG 420
Qy 421 NVDTPGSEDTMESRRSSMASTSTFTDTSIG---GPCRIRMLMLKRCMTRCRLILLI 477
Db 421 NVDTPGSEDTMESRRSSMASTSTFTDTSIG---GPCRIRMLMLKRCMTRCRLILLI 477
Qy 478 RLFRIMGIQISVYVSTIQHPPRDTTNGARLLGNPSAGIQSTYARLALSGGLRHDWGGLT 537
Db 478 RLFRIMGIQISVYVSTIQHPPRDTTNGARLLGNPSAGIQSTYARLALSGGLRHDWGGLT 537
Qy 480 QNM---GNTDSVVYSTIQHPPRDTTNGARLLGNPSAGIQSTYARLALSGGLRHDWGGLT 536
Db 480 QNM---GNTDSVVYSTIQHPPRDTTNGARLLGNPSAGIQSTYARLALSGGLRHDWGGLT 536
Qy 538 GGSNSAVNTSNPPAPGSHRFV 559
Db 538 GGSNSAVNTSNPPAPGSHRFV 559
Qy 537 GGSNSAVNTSNPPAPGSHRFV 558
Db 537 GGSNSAVNTSNPPAPGSHRFV 558

```

```

RESULT 3
Q47014 PRELIMINARY; PRT; 538 AA.
ID Q47014
AC Q47014
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE TRANSLOCATED intimin receptor TIR (EspE protein).
GN TIR OR ESPE.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=REPEC 84/110/1, and E65/56;
RA Krejany E.O.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=413/89-1;
RA Deibel C., Kraemer S., Chakraborty T., Ebel F.;
RX MEDLINE=96294040; PubMed=9632251;
RT "EspE, a novel secreted protein of attaching and effacing bacteria, is
RT directly translocated into infected host cells, where it appears as a
RT tyrosine-phosphorylated 90 kDa protein.";
RL Mol. Microbiol. 28:463-474 (1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=952G1;
RX MEDLINE=99003184; PubMed=9784578;
RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT "Translocated intimin receptors (Tir) of Shiga-toxicogenic Escherichia
RT coli isolates belonging to serogroups O26, O111, and O157 react with
RT sera from patients with hemolytic-uremic syndrome and exhibit marked
RT sequence heterogeneity.";
RL Infect. Immun. 66:5580-5586 (1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B10;
RA Marches O., Nougayrede J.-P., Boury M., Mainil J., Charlier G.,
RA Boullier S., De Rycke J., Milon A., Oswald E.;
RT "Role of Tir and Intimin in the pathogenesis of rabbit
RT enteropathogenic Escherichia coli.";
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=413/89-1;
RA Benkel P., Chakraborty T.;
RT "Genetic organisation and sequence of the LEE II locus in Shiga toxin-
RT producing Escherichia coli.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59502; AAC32028.2; -
DR EMBL; AJ223063; CAAL1065.1; -
DR EMBL; AF070068; AAC69316.1; -
DR EMBL; AF132728; AAD27868.1; -
DR EMBL; AF113597; AAF03080.1; -
DR EMBL; AJ277443; CAC81869.1; -
DR InterPro; IPR003536; Tir receptor.
DR Pfam; PF03549; Tir receptor.
DR PRINTS; PR01370; TRNSINTMINR.
KW Hypothetical protein; Receptor.
SQ SEQUENCE 538 AA; 55420 MW; 31D7A9E227B3D06C CRC64;

Query Match 62.2%; Score 1774.5; DB 2; Length 538;
Best Local Similarity 63.2%; Pred. No. 2.8e-89;
Matches 360; Conservative 61; Mismatches 106; Indels 43; Gaps 6;

Qy 1 MPIGNLGNPNVNNISIPAPLPQTDGAGG-RQGLINSTGPGSRALFTFVRNSMADSG 59
Db 1 MPIGNLGNPNVRALIPAPLPQTDGAGGARRQLNSNGPMGSRLLFTPIRNSVADAA 60

```

RC STRAIN-RDEC-1;
 RX MEDLINE=97055784; PubMed=8900070;
 RA Agin T.S., Cantey J.R., Boedeker E.C., Wolf M.K.;
 RT "Characterization of the eaeA gene from rabbit enteropathogenic
 Escherichia coli strain RDEC-1 and comparison to other eaeA genes from
 bacteria that cause attaching-effacing lesions."; PEMS Microbiol. Lett. 144:249-258(1996).
 RL [6]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-RDEC-1;
 RX MEDLINE=98254123; PubMed=9593291;
 RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
 RL Lai L.C., McNamee B.P., Sonnenberg M.S., Kaper J.B.;
 RT "The complete sequence of the locus of enterocyte effacement (LEE)
 from enteropathogenic Escherichia coli E2348/69."; Mol. Microbiol. 28:1-4(1998).
 RL [7]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-RDEC-1;
 RX MEDLINE=21153569; PubMed=11254564;
 RA Zhu C., Agin T.S., Elliott S.J., Johnson L.A., Thate T.E., Kaper J.B.,
 RL Boedeker E.C.;
 RT "Complete Nucleotide Sequence and Analysis of the Locus of Enterocyte
 Effacement from Rabbit Diarrheagenic Escherichia coli RDEC-1."; Infect. Immun. 69:2107-2115(2001).
 RL [8]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN-RDEC-1;
 RX MEDLINE=83/39;
 RA Boedeker E.C., Zhu C., Elliott S.J., Tonia T.S., Johnson L.A.,
 RL Thate T.E., Kaper J.B.;
 RT Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RL [9]
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=83/39;
 RX Tauschek M., Strugnell R.A., Robins-Browne R.M.;
 RA "Characterization of the LEE pathogenicity islands of rabbit
 enteropathogenic Escherichia coli."; Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RL EMBL; U59504; AAD19750.1; -;
 DR EMBL; AF045568; AAC15683.1; -;
 DR EMBL; U59503; AAB02941.1; -;
 DR EMBL; AF200363; AAK26722.1; -;
 DR EMBL; AF453441; AAL57549.1; -;
 DR InterPro; IPR003536; Tir_receptor.
 DR Pfam; PF03549; Tir_receptor; 1.
 DR PRINTS; PRO1370; TRNSINTIMNR.
 KW Receptor.
 SQ SEQUENCE 538 AA; 55410 MW; 3132A969B7B3D06C CRC64;

Query Match 62.0%; Score 1766.5; DB 2; Length 538;
 Best Local Similarity 63.0%; Pred. No. 7.8e-89;
 Matches 359; Conservative 61; Mismatches 107; Indels 43; Gaps 6;

QY 1 MPIGNLGNPNVNNISIPAPPPLPSQTDGAGG-RGQLINSTGLSGRALFTPVNRSMADSG 59
 DB 1 MPIGNLGNPNVNNISIPAPPPLPSQTDGAGG-RGQLINSTGLSGRALFTPVNRSMADSG 60
 QY 60 DNRAADVPGLPNMPLAASEITLNDGFVLDHGLDPLTNRIQIGSSVFRVETQDGGKH 119
 DB 61 DSRASDIPGLPTNPLFAAASEVSLHGALEVLHDHKGGLDPLNSAIGSSLFVRVETRDGSHV 120
 QY 120 AVCGRNGVETSVLSDOEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRI 179
 DB 121 AIGQKNGLETTVVLSSQFSSLSQSLDPEGKKNKFTVFTGGRGGAGHAMVTVASDIAEARQRI 180
 QY 180 LELLEPKGTGSGKAGSGKGVGLRESNSCAENTTETQTSTSTSLRSDPKLWALGTVA 239
 DB 181 IDKLEPKDTKETKEPG-----DPNSGEGKIIIEHTSTSTSLRADPKLWLSLGTIA 231
 QY 240 TGLIGLAATGIVQALALTPEPDSPTTDDPAAASATETATRDLTKEAFQNDPNQKNVID 299
 DB 232 AGLIGNAATGIAQVALTPEPDPITTDPAANAATAAQAQDLTKKEAFQNDPNQKNVID 291

QY 300 ELGNAIPSGVLKDDVVANIIEQAKAAGEAKQQAQIENNAQAQKYQDEQAQKQEBELKVSS 359
 DB 292 ENGNAIPSGELKDDVVAQIAEQAKAAGEAQAEIESNSQAQKYQDEQAKREQEMSLSS 351
 QY 360 GAGYGLSGALILGGGIGVAVTAALHRKNQPVQVOTTTTTTTTTTSARTVENKPNNTPAQ 419
 DB 352 GVGYGISGALILGGGIGAGVTAALHRKNQPAEQTTTTRT-----VVDNQPTNNASAQ 403
 QY 420 GNVDTPEGSEDTMESRRESSMASTSTFFDTSS-----IGGPCRIRMLMLKHKRCMIR 469
 DB 404 GNTDTSGPESPASRRNSNASLASNGSDTSSGTGVENPYADVGMFPRNDSLARISEPIYD 463
 QY 470 RCRLLILIRLFRWIGIQISVYVSTIHOHPRDPTDNGARLGNPSAGIQSTYARLALSGGL 529
 DB 464 E-----VAADPNYSVIQHFSGNSPVTG-RLVGTGPGQIQSTYALLASSGGL 508
 QY 530 RHDMGGLTGGSNSAVNTSNPPAPGSHRFV 559
 DB 509 RLGMGGLTGGSASVSTANAAPTGPAPRFV 538

RESULT 5
 O85508 PRELIMINARY; PRT; 538 AA.
 ID O85508;
 AC O85508;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE Translocated intimin receptor Tir.
 GN Tir.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Escherichia.
 OX NCBI_TaxID:562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EPEC87A;
 RX MEDLINE=99003184; PubMed=9784578;
 RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
 RT "Translocated intimin receptors (Tir) of shiga-toxicogenic Escherichia
 coli isolates belonging to serogroups O26, O111, and O157 react with
 sera from patients with hemolytic-uremic syndrome and exhibit marked
 sequence heterogeneity."; Infect. Immun. 66:5580-5586(1998).
 RL EMBL; AF070069; AAC69318.1; -;
 DR InterPro; IPR003536; Tir_receptor.
 DR Pfam; PF03549; Tir_receptor; 1.
 DR PRINTS; PRO1370; TRNSINTIMNR.
 KW Receptor.
 SQ SEQUENCE 538 AA; 55602 MW; 447052A0E3214D6D CRC64;

Query Match 61.5%; Score 1753.5; DB 2; Length 538;
 Best Local Similarity 62.6%; Pred. No. 4e-88;
 Matches 357; Conservative 61; Mismatches 109; Indels 43; Gaps 6;

QY 1 MPIGNLGNPNVNNISIPAPPPLPSQTDGAGG-RGQLINSTGLSGRALFTPVNRSMADSG 59
 DB 1 MPIGNLGNPNVNNISIPAPPPLPSQTDGAGG-RGQLINSTGLSGRALFTPVNRSMADSG 60
 QY 60 DNRAADVPGLPNMPLAASEITLNDGFVLDHGLDPLTNRIQIGSSVFRVETQDGGKH 119
 DB 61 DSRASDIPGLPTNPLFAAASEVSLHGALEVLHDHKGGLDPLNSAIGSSLFVRVETRDGSHV 120
 QY 120 AVCGRNGVETSVLSDOEYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRI 179
 DB 121 AIGQKNGLETTVVLSSQFSSLSQSLDPEGKKNKFTVFTGGRGGAGHAMVTVASDIAEARQRI 180
 QY 180 LELLEPKGTGSGKAGSGKGVGLRESNSCAENTTETQTSTSTSLRSDPKLWALGTVA 239
 DB 181 IDKLEPKDTKETKEPG-----DPNSGEGKIIIEHTSTSTSLRADPKLWLSLGTIA 231
 QY 240 TGLIGLAATGIVQALALTPEPDSPTTDDPAAASATETATRDLTKEAFQNDPNQKNVID 299


```
Db 232 AGLGMAATGTAQAVALTPEDDTTTDDPTTAATAEAATKDRLTQBAFQDDPKQKNVID 291
Qy 300 ELGNAPSGVLKDDVWVIAEQAQAAGAEAKQAQAIENNAQAQKKYDEQAQAKROBELKYSS 359
Db 292 ENGNAIPSGELIDVVAQIAEQAKAQAQAEIAESQAQKKYDEQAQAKROBELKYSS 351
Qy 360 GAGVGLSGLILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTTTSAKTVEKNPANTPAQ 419
Db 352 GVGVGISGALILGGGIGAGVTAALHRKNQPAEQITITRT-----VVDNQPTNNSAQ 403
Qy 420 GNVDTPGSEDTWESRRSSMASTSTFFDTSS-----IGGPCRIRMLMKHRCMIR 469
Db 404 GNTDTSGPEEPAGRRNSNAGLSNGSDTSSTGVENPYADVPRNDLSLARIPEEPIYD 463
Qy 470 RCRLLILRLFRWIGIQISVVYSTIQHPRTDTTNGARLLGNPSAGIQSTYARLALSGGL 529
Db 464 E-----VAADPNYSVIQHPSGNSPVTVG-RLVGTGPGGIGQSTYALLASSGGL 508
Qy 530 RHDGGLTGGNSAVNTSNPPAPGSHRFV 559
Db 509 RLCMGGLTGGGSAVSTANASPTGPARFV 538

RESULT 6
O68258
ID O68258 PRELIMINARY; PRT; 551 AA.
AC O68258
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Translocated intimin receptor.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98187918; PubMed=9529069;
RA Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT "Molecular analysis of Shiga toxicogenic Escherichia coli O111:H-
RT proteins which react with sera from patients with hemolytic-uremic
RT syndrome.";
RL Infect. Immun. 66:1467-1472 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95003184; PubMed=9784578;
RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT "Translocated intimin receptors (Tir) of shiga-toxicogenic Escherichia
RT coli isolates belonging to serogroups O26, O111, and O157 react with
RT sera from patients with hemolytic-uremic syndrome and exhibit marked
RT sequence heterogeneity.";
RL Infect. Immun. 66:5580-5586 (1998).
DR EMBL; AF025311; AAC69249.1; -.
DR InterPro; IPR003536; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTMINR.
KW Receptor.
SQ SEQUENCE 551 AA; 56975 MW; 6EC95F76EF0F44CC CRC64;

Query Match 55.1%; Score 1570; DB 2; Length 551;
Best Local Similarity 59.5%; Pred. No. 4.4e-78;
Matches 339; Conservative 67; Mismatches 134; Indels 30; Gaps 11;

Qy 1 MPIGNLGNPNVNSIAPPPLPSQTDGA--GGRGQLINSTGPGSLRALFTPVNRNSMADS 58
Db 1 MPIGNLGNPNVNSNLIAPPPLPSQTDGAGRGAGQLINSTGALGSLRLLFPRLNSIADS 60
Qy 59 GDNRASDPVGLPVNPMRL--AASEITLNDGFEVLHDHGFLDTLNKQIGSSVFRVETQBDG 116
Db 61 VDSR--DIPGLPVNPSRLAATSETCLLGGFEVLHDHGFLDTLNKQIGASAFRIEQQSDG 118
Qy 117 KHIAVGQRNGVETSVVLSDQYARLQSIDPBGKDFVFTGGRGGAGHAMVTVASDITEAR 176
```

```
Db 119 SYAAIGKNGVEVSVILNSQELQDAIDIEDKGRFVFTGGRGGGSHMTVPASDIABAR 178
Qy 177 QRILLELEPKGTGSKAG---ESKGVGELRESNSGAENT--TETQTSTSTSSLRSDPKL 231
Db 179 AKILAKLDPNNHGGSOARNVDTRSVGVG-----SASGMDSDSVSETRTSSTASSVRSDFK 234
Qy 232 WLALGTVATGLIGLAATGIVQALALTPEPDSPTTTDDPDAASATETATRDOLTKEAFONP 291
Db 235 WVSIGATAAGLAGLAATGITOALALTPEPDDPTTTDPEQAASAAESATRDOLTQEAFFNP 294
Qy 292 DNQKVNIDELGNATPSGVLDKDDVVANIEQAQAAGEAKQAQAIENNAQAQKKYDQQAQR 351
Db 295 ENQKVSIDEIGNSIPSGELKDDVVAKIEQAQAAGEAKQAQAVESNAQAQQRIDTQYARR 354
Qy 352 QEELKVSAGAGYLSGALILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTTTSAKTVEKN 411
Db 355 QEELKVSAGAGYLSGALILGGGIGVAVTAALHRKNQPAEQITITRT-----VVOQQ 408
Qy 412 PANTPAQGNVDTPGSEDTWESRRSSMASTSTFF-DTSSIGGPCRIRMLMKHRCMIR 470
Db 409 TCGNTPAQGGTDAIRAEDTSLNRRDSQRSTASTHWSDTSSAVVNPYAEVGEARNSPARQ 468
Qy 471 CRLILRLFRWIGIQISVVYSTIQHPRTDTTNGARLLGNPSAGIQSTYARLA--LSGGL 529
Db 469 ABEHIYDE-----VAADPNYSVIQHPSGNNQVTG-RLMGTPGGIGQSTYAILTNNSAGL 521
Qy 530 RHDGGLTGGNSAVNTSNPPAPGSHRFV 559
Db 522 RLCMGGLTGGGSAVNTANAPTPGPARFV 551

RESULT 7
O9KWH9
ID O9KWH9 PRELIMINARY; PRT; 552 AA.
AC O9KWH9
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tir.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HK01;
RA Abe A., Nagano H.;
RT "Analyses of type III secreted proteins and Tir in enteropathogenic
RT Escherichia coli O157:H45.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB036053; BAA96815.1; -.
DR InterPro; IPR003536; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTMINR.
SQ SEQUENCE 552 AA; 57005 MW; A8D79EE22EE50A4B CRC64;

Query Match 54.5%; Score 1554.5; DB 2; Length 552;
Best Local Similarity 57.5%; Pred. No. 3.1e-77;
Matches 336; Conservative 63; Mismatches 128; Indels 57; Gaps 12;

Qy 1 MPIGNLGNPNVNSIAPPPLPSQTDGA--GGRGQLINSTGPGSLRALFTPVNRNSMADS 58
Db 1 MPIGNLGNPNVNSNLIAPPPLPSQTDGARGGTGHLISSTGALGSLRLLFPRLNSVADS 60
Qy 59 GDNRASDPVGLPVNPMRLAA--SEITLNDGFEVLHDHGFLDTLNKQIGSSVFRVETQBDG 116
Db 61 VDSR--DIPGLPANPSRLAATSETCLLGGFEVLHDHGFLDTLNQIQSPSAFVEAQADG 118
Qy 117 KHIAVGQRNGVETSVVLSDQYARLQSIDPBGKDFVFTGGRGGAGHAMVTVASDITEAR 176
Db 119 THAAIGKNGLEVSVLSPQELQSIDIEGKNRFTVFTGGRGGSGHPMTVVASDIABAR 178
```



```

Db      1 MPIGNLGNHNPNVNSIPAPPLPSQTDGA--GGRGQLINSTGPIGSRLLFPSPRSSIVDT 60
Qy      59 GDNRASDPVGLPVNPMRLAASEITLNDGFEVLHGHGPDLTINRQIGSSVFRVETQEDGKH 118
Db      61 VDSR--DVPGLPEPLRATSETCLHGGFEVLHGHGPDLTINRQIGSSVFRVETQEDGKH 118
Qy      119 IAVGQRNGVTSVVLSDQYARLOSIDPEGKDKFVFTGGRGAGHAMVTVASDIPEAROR 178
Db      119 AAGVKGQGVSVTLNSELQSLDTEGTRFVFTGGRGSGHAMVTVASDISQAREK 178
Qy      179 ILELLEPKGTG-----ESKGAGESKGVGELRESNGAENTTETOTSTSTSLRSDP 229
Db      179 IIAKLDPNHHGRQPKDIDTRSVGVGSAGMGD-----GV--VSETHSTTTSSVRSDP 230
Qy      230 KLWLAGTVAATGLGLAATGIVQALALTPPEPDSPTTTDPDAAASATETATRDQLTKEAFQ 289
Db      231 KFWSVGAGIAGLAGLAATGIVQALVTPAPDDPTTTPDDEAANAAREATKDQLTKEAFQ 290
Qy      290 NPDNQKNIDELGNAIPSGVLKDDVVANIEBQAKAAGEEAKQQAIIENNAQAKKYDEQQA 349
Db      291 NPDNQKNIDELGNAIPSGELKDDVVAQIADQAKVAGEQARQAVESNAQAQRHDDQQA 350
Qy      350 KQBELKVSSGAGVGLSGALILGGIGVAVTAALHRKNQPVETQTTTTTTTTTSARTVE 409
Db      351 KQQLDLSGGIGVGLSALIVGGIGAGVTAMLHRRNPPEQTITATTHS-----VIQ 403
Qy      410 NKPAANTPAQGNVDTPGSEDTMESRRSSMASTST--PFDTS--SIGGPCRIRMLMKHRCM 467
Db      404 QOTGNTTRAQGAADTTGVENASLFRDSQASVASTQWSDTSGDVVNP-----YAEQWM 456
Qy      468 IRRCLLILRLFRWIGQISVYSTIQHPPRDITDNGARLLGNPSAGIQSTYARLALSG 527
Db      457 SRNPSLLAPEPIYDEVAPDPNYSVIOHFSGNPNVTG--RLVGSFGQGIQSTYALLASSG 515
Qy      528 GLRDMGGLTGGSSNAVNTSNPPAPGSHRFV 559
Db      516 GLRLGMGLTGGESAGSANAATTPGVERFV 547

RESULT 10
Q9ET11 ID Q9ET11 PRELIMINARY; PRT; 547 AA.
AC Q9ET11;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Translocated intimin receptor Tir.
OS Citrobacter rodentium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX Enterobacteriaceae; Citrobacter.
RN [1]_TaxID=67825;
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1843-73T, and DBS100;
RA MEDLINE=20553330; PubMed=11101562;
RX Luperchio S.A., Newman J.V., Dangler C.A., Schrenzel M.D.,
RA Brenner D.J., Steigerwalt A.G., Schauer D.B.;
RT "Citrobacter rodentium, the Causative Agent of Transmissible Murine
RT Colonic Hyperplasia, Exhibits Clonality: Synonymy of C. rodentium and
RT Mouse-Pathogenic Escherichia coli.";
RL J. Clin. Microbiol. 38:4343-4350(2000).
DR EMBL; AF301618; AAG40758.1; -
DR EMBL; AF301617; AAG25642.1; -
DR InterPro; IPR003536; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTIMINR.
KW Receptor.
SQ SEQUENCE 547 AA; 56270 MW; CB8318B301049C37 CRC64;

```

Query Match 53.3%; Score 1521; DB 2; Length 547;
 Best Local Similarity 57.2%; Pred. No. 2.1e-75;
 Matches 327; Conservative 68; Mismatches 139; Indels 38; Gaps 10;

```

Qy      1 MPIGNLGNHNPNVNSIPAPPLPSQTDGA--GGRGQLINSTGPIGSRLLFPSPRSSIVDT 58
Db      1 MPIGNLGNHNPNVNSIPAPPLPSQTDGATRCNGSSLLISSTGSLGRLLFPSPRSSIVDT 60
Qy      59 GDNRASDPVGLPVNPMRLAASEITLNDGFEVLHGHGPDLTINRQIGSSVFRVETQEDGKH 118
Db      61 VDSR--DVPGLPEPLRATSETCLHGGFEVLHGHGPDLTINRQIGSSVFRVETQEDGKH 118
Qy      119 IAVGQRNGVTSVVLSDQYARLOSIDPEGKDKFVFTGGRGAGHAMVTVASDIPEAROR 178
Db      119 AAGVKGQGVSVTLNSELQSLDTEGTRFVFTGGRGSGHAMVTVASDISQAREK 178
Qy      179 ILELLEPKGTG-----ESKGAGESKGVGELRESNGAENTTETOTSTSTSLRSDP 229
Db      179 IIAKLDPNHHGRQPKDIDTRSVGVGSAGMGD-----GV--VSETHSTTTSSVRSDP 230
Qy      230 KLWLAGTVAATGLGLAATGIVQALALTPPEPDSPTTTDPDAAASATETATRDQLTKEAFQ 289
Db      231 KFWSVGAGIAGLAGLAATGIVQALVTPAPDDPTTTPDDEAANAAREATKDQLTKEAFQ 290
Qy      290 NPDNQKNIDELGNAIPSGVLKDDVVANIEBQAKAAGEEAKQQAIIENNAQAKKYDEQQA 349
Db      291 NPDNQKNIDELGNAIPSGELKDDVVAQIADQAKVAGEQARQAVESNAQAQRHDDQQA 350
Qy      350 KQBELKVSSGAGVGLSGALILGGIGVAVTAALHRKNQPVETQTTTTTTTTTSARTVE 409
Db      351 KQQLDLSGGIGVGLSALIVGGIGAGVTAMLHRRNPPEQTITATTHS-----VIQ 403
Qy      410 NKPAANTPAQGNVDTPGSEDTMESRRSSMASTST--PFDTS--SIGGPCRIRMLMKHRCM 467
Db      404 QOTGNTTRAQGAADTTGVENASLFRDSQASVASTQWSDTSGDVVNP-----YAEQWM 456
Qy      468 IRRCLLILRLFRWIGQISVYSTIQHPPRDITDNGARLLGNPSAGIQSTYARLALSG 527
Db      457 SRNPSLLAPEPIYDEVAPDPNYSVIOHFSGNPNVTG--RLVGSFGQGIQSTYALLASSG 515
Qy      528 GLRDMGGLTGGSSNAVNTSNPPAPGSHRFV 559
Db      516 GLRLGMGLTGGESAGSANAATTPGVERFV 547

RESULT 11
O50190 ID O50190 PRELIMINARY; PRT; 549 AA.
AC O50190;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Translocated intimin receptor.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX Enterobacteriaceae; Escherichia.
RN [1]_TaxID=562;
RP SEQUENCE FROM N.A.
RC STRAIN=E2348/59;
RX MEDLINE=98050926; PubMed=9390560;
RA Kenny B., Devinney R., Stein M., Reinscheid D.J., Frey E.A.,
RA Finlay B.B.;
RT "Enteropathogenic E. coli (EPEC) transfers its receptor for intimate
RT adherence into mammalian cells.";
RL Cell 91:511-520(1997).
DR EMBL; AF013122; AAB88410.1; -
DR InterPro; IPR003536; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTIMINR.
KW Receptor.
SQ SEQUENCE 549 AA; 56843 MW; 40C8B8B234409A08 CRC64;

```

Query Match 52.4%; Score 1495; DB 2; Length 549;
 Best Local Similarity 54.7%; Pred. No. 5.5e-74;
 Matches 322; Conservative 65; Mismatches 132; Indels 70; Gaps 12;

Qy 1 MPIGNLGNPNVNNISIPAPPLPSOTDGA--GGRGQLNSGPGLSRALFTEVNSWADS 58
 Db 1 MPIGNLGNVNGNHLIPAPPLPSOTDGAARGGTGHLLISGALGSLFSPKLNWADS 60
 Qy 59 GDNRASDPVGLPVPNMLAA--SEITLNDGFEVLHDHGPLDNLNRQIGSSVFRVETOEDG 116
 Db 61 VDSR--DIPGLPTWPSRLAAATSETCLLGGFEVLHDHGKPLDNLNTQIGPSAFRVEVQADG 118
 Qy 117 KHIAVGQRNGVETSIVLSQDQBYARLQSIDPBGDKFVFTGGRGAGHAMVTVASDITEAR 176
 Db 119 THAAIGKNGLEVSTLSPQWSSLSQIDTEGKNRFVFTGGRGSGHPMTVVASDIAEAR 178
 Qy 177 QRILELLEPKGTG-----ESKGAGESKGVCELESNGSAENTTETQTSTSTSLRS 227
 Db 179 TKILAKLPDNDHGRQPKVDVTRSVGVGSASGI-----DDGV--VSETHFTSTNVSRS 230
 Qy 228 DPKLWALGTWATGLIGLAAGTGIQVALALTEPDPSTPTTDPDAASATETATRDQLTKEA 287
 Db 231 DPKFVWSVGAIAAGLAGLAATGIAQALALTEPDDPTTDDQAANAESATKQDLTQEA 290
 Qy 288 FQNPQNKVNDLGNALPSGVLKDDVVANTEEQAAGEAKQAQIENNAQAOKYDEQ 347
 Db 291 FQNPENQKVNIDANGNALPSGELKDDIVEQIAQAQAGEVARQQAQVESAQAQRYEDQ 350
 Qy 348 QAKRQBELKSVGAGYGLSGALILGGGIGVATRALHKNOPVETTTTTTTTTTSART 407
 Db 351 HARRQBELQSLGGIGYGLSSALIVAGGIGAGVTTALHRRNOPAEQTTTTTT-----HT 403
 Qy 408 VENKPNANTPAQGNVDTPGSEDMSRSMSTSTFTDSS-----IGPCPRIR 458
 Db 404 VVQQTGGIPQHKVAMQERRRFDSDRDSQGSVASTHWSDDSSSVNPNVAEYVG--AR 460
 Qy 459 MMLKHCRCMIRCLLILRLFRWIGIQISVVYSTIOHPP-----RDTTDNG---ARLLG 510
 Db 461 NLSLSAHQ-----PEEHIYDEVAADPGVSVIQNFGSGGVPVTRGLIG 500
 Qy 511 NPSAGTQSVARLALSGGLRHDWGLTGSSNVAVNTSNPPAPGSHRFV 559
 Db 501 TPGQIGTQSVALLANSGLGLGMLGGLTSGGETAVSVNAAPFGPGRFV 549
 RESULT 12
 Q9V163 PRELIMINARY; .PRT; 1323 AA.
 AC Q9V163 Q9NHX4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE CG1070 protein (Alhambra).
 GN ALHAMBRA OR CG1070.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer V.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Heit G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.P., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Borchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foslter C., Gabrielian A.E., Gang N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.W., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 [2]
 RN SEQUENCE FROM N.A.
 RP Celniker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M., D.,
 RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Garg N.S., George R.A.,
 RA Gonzaez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacle J.J., Paragass V., Park S., Patel S., Pfeiffer B., Scheeler F.,
 RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [3]
 RN SEQUENCE FROM N.A.
 RP Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
 RT "Annotation of Drosophila melanogaster genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE FROM N.A.
 RP Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 [5]
 RN SEQUENCE FROM N.A.
 RP FlyBase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 [6]
 RN SEQUENCE FROM N.A.
 RP Perrin L., Dura J.M.;
 RT "Alhambra, a Drosophila homolog of mammalian AF-10 and AF-17.";
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AE003672; AAF54062.2; --
 DR EMBL; AF217960; AAF72595.1; --
 DR FlyBase; FBN0037471; Alhambra.
 DR InterPro; IPR000104; Antifreeze_1.
 DR InterPro; IPR001965; Znf_PHD.
 DR Pfam; PF00628; PHD; 1.
 DR PRINTS; PR00308; ANTIREEZEI.

DR SMART; SM00249; PHD; 2.
DR PROSITE; PS50016; ZF_PHD 2; 1.
SQ SEQUENCE 1323 AA; 132884 MW; D53C0C8AF392F9A6 CRC64;

Query Match 6.7%; Score 190; DB 5; Length 1323;
Best Local Similarity 20.2%; Pred. No. 0.038;
Matches 131; Conservative 90; Mismatches 273; Indels 154; Gaps 23;

Qy 3 IGNLGNPNVNSIPPAPLPQSDTGAGRGQLINSTGLSRALFTFVRNSMADSGDNR 62
Db 322 IGNISN--SLNNLPGG---SSSTSSAG-----NVPGGSGGIISASSGGATQSTSSQ 368
Qy 63 ASDVGPLVPMRLAASEITLNDGFEVLHDGHLDTLNRQIGSSVFRVEQEDKHTAVG 122
Db 369 SSTAPGTTKSS---ASSSSSSNSYKEKHS-----KSLSKSTSSK-----DKDGKDSSTN 415
Qy 123 QRMGVET---SVVLSQDYARLQSIDPEGKDK-----F 152
Db 416 SANNNFTNSASSTSSNSSTREKSSKLSKNKDSNQVPSATSSLSSTSSINTOPSSSTS 475
Qy 153 VFTGGRGAG-HAMVTVASDITEARQRIE-----LLEPKGTGESKGAGSKGVGLRE 205
Db 476 TATAGSGGTGTHVSSSAAGLNSAPSTTNEHSNHAHTLSTNGTCAGSAAGKLGVSNLN 535
Qy 206 SNGG-----AENTTETQTSSTSLRSDPKLWALGTATGLIGLAA--TGIVQ 252
Db 536 SSSGFGDLRSVSTSSSTVNDSTGFGCSNSRENLSGAGSSASNMPGTIAPGTGGVS 595
Qy 253 ALALTPEDSPPTTDPDAASATATRDQITKEAFQNPQKVNIDELGNAIPSGVLKD 312
Db 596 SSAATNLSTNKGSSSTANSLTSTST-----SSGSSNSSSKKRKAD----- 638
Qy 313 DVVANIEEQAKAGE-EAKQOAIENNAQAOKKYD-----EQOAKRQE 353
Db 639 -----SAKSTSSISTSGSALDNNLSIRYDIKDQVQALPLTDPEKEIEKSKRQR 690
Qy 354 -ELKV---SSGAGVGLSGALILGGIGVAVTAALHKNQPVQETTTTTTTTSARTVE 409
Db 691 TELSPPTHQTSATAEVNAPLASSTASIAVTASATAASAPPATGTTTLGSSISGNAGSTS 750
Qy 410 NKPAANTPAQGNVDT-----PGSEDTMESRRSSMASTSTFPDTSISGPGCIRMLML 462
Db 751 SGNSSGVSATGGAQSAVSGGYPKTESKSSGTASAGSGSSNTSTKHGSIKDISS 810
Qy 463 KHRCMIRRCRLILIRLIRIWIQIISVYVSTIQHP-----PRDTTDNG-----ARLLGNPSA 514
Db 811 SSNOQASTASSSSAPSLY-----VSVPLSTANVPNGINLPSTSSSTTSSSHSASSRSS 864
Qy 515 GIQSTYARLALSGGLRHDMMGLTG-----GSNSAVNTSNPPAP 553
Db 865 GAQSOHQQLSNALVGPMPGGSAGAFHGGTTTSAGSSSVIQOQSGKSSP 912

RESULT 13
Q9BMZ9
ID Q9BMZ9 PRELIMINARY; PRT; 1376 AA.
AC Q9BMZ9,
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE AF10.
GN ALHAMBRA OR CG1070.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Linder B., Jackie H.;
RT "The Drosophila homolog of AF10 is expressed in a striped pattern
throughout blastoderm."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF281145; AAK06385.1; --
DR FlyBase; FBgn0037471; Alhambra.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR SMART; SM00249; PHD; 2.
DR PROSITE; PS50016; ZF_PHD 2; 1.
SQ SEQUENCE 1376 AA; 138906 MW; A40DF3CF5121B17C CRC64;

Query Match 6.7%; Score 190; DB 5; Length 1376;
Best Local Similarity 20.2%; Pred. No. 0.04;
Matches 131; Conservative 90; Mismatches 273; Indels 154; Gaps 23;

Qy 3 IGNLGNPNVNSIPPAPLPQSDTGAGRGQLINSTGLSRALFTFVRNSMADSGDNR 62
Db 375 IGNISN--SLNNLPGG---SSSTSSAG-----NVPGGSGGIISASSGGATQSTSSQ 421
Qy 63 ASDVGPLVPMRLAASEITLNDGFEVLHDGHLDTLNRQIGSSVFRVEQEDKHTAVG 122
Db 422 SSTAPGTTKSS---ASSSSSSNSYKEKHS-----KSLSKSTSSK-----DKDGKDSSTN 468
Qy 123 QRMGVET---SVVLSQDYARLQSIDPEGKDK-----F 152
Db 469 SANNNFTNSASSTSSNSSTREKSSKLSKNKDSNQVPSATSSLSSTSSINTOPSSSTS 528
Qy 153 VFTGGRGAG-HAMVTVASDITEARQRIE-----LLEPKGTGESKGAGSKGVGLRE 205
Db 529 TATAGSGGTGTHVSSSAAGLNSAPSTTNEHSNHAHTLSTNGTCAGSAAGKLGVSNLN 588
Qy 206 SNGG-----AENTTETQTSSTSLRSDPKLWALGTATGLIGLAA--TGIVQ 252
Db 589 SSSGFGDLRSVSTSSSTVNDSTGFGCSNSRENLSGAGSSASNMPGTIAPGTGGVS 648
Qy 253 ALALTPEDSPPTTDPDAASATATRDQITKEAFQNPQKVNIDELGNAIPSGVLKD 312
Db 649 SSAATNLSTNKGSSSTANSLTSTST-----SSGSSNSSSKKRKAD----- 691
Qy 313 DVVANIEEQAKAGE-EAKQOAIENNAQAOKKYD-----EQOAKRQE 353
Db 692 -----SAKSTSSISTSGSALDNNLSIRYDIKDQVQALPLTDPEKEIEKSKRQR 743
Qy 354 -ELKV---SSGAGVGLSGALILGGIGVAVTAALHKNQPVQETTTTTTTTSARTVE 409
Db 744 TELSPPTHQTSATAEVNAPLASSTASIAVTASATAASAPPATGTTTLGSSISGNAGSTS 803
Qy 410 NKPAANTPAQGNVDT-----PGSEDTMESRRSSMASTSTFPDTSISGPGCIRMLML 462
Db 804 SGNSSGVSATGGAQSAVSGGYPKTESKSSGTASAGSGSSNTSTKHGSIKDISS 863
Qy 463 KHRCMIRRCRLILIRLIRIWIQIISVYVSTIQHP-----PRDTTDNG-----ARLLGNPSA 514
Db 864 SSNOQASTASSSSAPSLY-----VSVPLSTANVPNGINLPSTSSSTTSSSHSASSRSS 917
Qy 515 GIQSTYARLALSGGLRHDMMGLTG-----GSNSAVNTSNPPAP 553
Db 918 GAQSOHQQLSNALVGPMPGGSAGAFHGGTTTSAGSSSVIQOQSGKSSP 965

RESULT 14
Q962D1
ID Q962D1 PRELIMINARY; PRT; 1321 AA.
AC Q962D1,
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Zinc finger/leucine zipper protein DALF isoform C3.
GN ALHAMBRA OR CG1070.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 21.694 Seconds
(without alignments)
1211.758 Million cell updates/sec

Title: US-09-189-415A-4

Perfect score: 2851

Sequence: 1 MPIGNLGHNPVNNNSIPPAP.....SNSAVNTSNNPPAPGSHRFV 559

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	146.5	5.1	1120	1 STFR_ECOLI	P76072 escherichia
2	142	5.0	563	1 MUCS_BOVIN	P98091 bos taurus
3	140.5	4.9	1026	1 STAU_DROME	P25159 drosophila
4	136.5	4.8	461	1 US45_LACLC	P22865 lactococcus
5	133.5	4.7	3178	1 YS99_CAEEL	O09624 caenorhabdi
6	133	4.7	439	1 Y579_CHLTP	O84583 chlamydia t
7	132.5	4.6	755	1 Y572_CHLPN	Q92791 chlamydia p
8	132.5	4.6	1258	1 ICEN_ERWHE	P16239 erwinia her
9	131	4.6	2660	1 YEEJ_ECO57	O8x8v7 escherichia
10	130.5	4.6	1034	1 ICEN_PANAN	Q47879 pantoea ana
11	130	4.6	1077	1 HLES_DROME	Q02308 drosophila
12	129.5	4.5	2411	1 DAB_DROME	P98081 drosophila
13	129	4.5	666	1 FLID_VIBCH	Q9Kq63 vibrio chol
14	129	4.5	784	1 SP4_HUMAN	Q02446 homo sapien
15	129	4.5	1460	1 PMPC_CHLMU	Q9pjy1 chlamydia m
16	129	4.5	2090	1 HFC1_MESAU	P51611 mesocricetu
17	128.5	4.5	1902	1 P3P_LACLC	P15292 lactococcus
18	128	4.5	5703	1 MUSE_HUMAN	Q9hc84 homo sapien
19	127.5	4.5	436	1 Y868_CHLMU	Q9pj91 chlamydia m
20	127.5	4.5	3726	1 TRX_DROME	P20659 drosophila
21	127	4.5	1337	1 DEXT_STRDO	P39653 streptococ
22	126	4.4	997	1 BIR1_SCHPO	O14064 schizosacch
23	125.5	4.4	1322	1 ICEA_PANAN	P20469 pantoea ana
24	125	4.4	1694	1 IGA0_HAEIN	P44969 haemophilus
25	125	4.4	1702	1 IGA2_HAEIN	P45384 haemophilus
26	125	4.4	1772	1 MSP1_PLAYO	P13828 plasmodium
27	125	4.4	1802	1 HKR1_YEAST	P41809 saccharomyc
28	125	4.4	1902	1 P1P_LACLC	P16271 lactococcus
29	124	4.3	1102	1 YG45_SCHPO	O60184 schizosacch
30	124	4.3	1902	1 P2P_LACLC	P15293 lactococcus
31	123.5	4.3	2483	1 PCX_DROME	P18490 drosophila
32	123	4.3	782	1 SP4_MOUSE	Q62445 mus musculu
33	123	4.3	1007	1 Y741_CHLMU	Q9pj16 chlamydia m

34 123 4.3 1849 1 IGA4_HAEIN P45386 haemophilus
35 122.5 4.3 2842 1 APC_RAT P70478 rattus norv
36 122 4.3 2003 1 YDBA_ECOLI P33666 escherichia
37 122 4.3 2035 1 HFC1_HUMAN P51610 homo sapien
38 121.5 4.3 917 1 SMOO_HUMAN P53814 homo sapien
39 121.5 4.3 1079 1 IF2P_SCHPO Q10251 schizosacch
40 121.5 4.3 1189 1 YJH6_YEAST P70735 saccharomyc
41 121 4.2 1150 1 AFPMU_PIG P12021 sus scrofa
42 121 4.2 1306 1 MSB2_YEAST P32334 saccharomyc
43 120.5 4.2 746 1 7UP2_DROME P16376 drosophila
44 120 4.2 797 1 VGLX_HSVBE P28968 equine herp
45 120 4.2 1508 1 BCSC_XANAC P58938 xanthomonas

ALIGNMENTS

RESULT 1
STFR_ECOLI
ID STFR_ECOLI STANDARD; PRT; 1120 AA.
AC P76072; P77560;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Side tail fiber protein homolog from lambdoid prophage Rac.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
EX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R.; Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
EX MEDLINE=97251357; PubMed=9097039;
RA Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K., Itoh T.,
RA Kasai H., Kashimoto K., Kimura S., Kitagawa M., Kitagawa M.,
RA Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA Sampei G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 28.0-40.1 min region on the linkage map.";
RL DNA Res. 3:363-377(1996).
CC -!- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000234; AAC74454.1; ALT_INIT.
CC EMBL; D90774; BAA14966.1; -.
CC EMBL; D90775; BAA14975.1; -.
CC PIR; G64887; G64887.
CC EcoGene; EG13370; stfr.
CC InterPro; IPR005003; Phage fiber.
CC InterPro; IPR005068; Phage fiber_2.
CC Pfam; PF03335; Phage fiber_6.
CC Pfam; PF03406; Phage fiber_2; 1.
KW Hypothetical protein; fiber protein; Repeat; Complete proteome.


```
SQ SEQUENCE 1120 AA; 113779 MW; 542E59D71BE795B4 CRC64;
Query Match 5.1%; Score 146.5; DB 1; Length 1120;
Best Local Similarity 23.1%; Pred. No. 0.46;
Matches 92; Conservative 52; Mismatches 163; Indels 91; Gaps 17;

QY 79 SEITLNDGFEVLH-----DHGFLDNLNRQIGSSVFRVETQEDGKHIAVGQ-----123
DB 59 SVILLVEGPPPSHAGTITVYEDSQP-GTLNDFLGM-----TEDDARPEALRRPELMVEE 112
QY 124 --RNGVETSVVLSDOEYARLQSIDPEGDKFVFTGGRGGAGHAMVTVASDITEARQRIE 181
DB 113 VARN--ASAVAQNTAAKKSASDAS-----TSAREAAATHA--ADAADGARA-----154
QY 182 LLEPKGTGSKGAGSKGVGELRENSGAENTTETQTSSTSLSDPKLMLALGTATG 241
DB 155 -----ASTSAGQAASSAQAASSAGTASTKATEASAKAAESSK-----SAAAT 199
QY 242 LIGLAATGIVQALA---LTPEPDSPTTTPDPAASATETATRDQLTKEAFONPDNQVN 297
DB 200 SAGAARTSETNASLSQSAATSAATTKASEATTSARDAA-----SKEAKSSET---251
QY 298 IDELGNAPISGVKDDVNVANIEQAKAGEEAK-QQAIENNAQAQKYDEQQAQKQEBLK 356
DB 252 -----NASSS-----ASSAASSATAAGNSAKAATSETNARSSETAAGQASAAAGSK 299
QY 357 VSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVQTITTTTTTTTSARTVENKPNANT 416
DB 300 ---TAAASSAASASTSAGQASATATA---GKSAESAASSASTATTAKAGATEQASAAAR 353
QY 417 PAQ----GNVDTPGSDTWMESRRSSMASTSTFFDTSS 450
DB 354 SASAAKTSNNAKASSETSSESKTAASASSAASSAS 391

RESULT 2
MUCS_BOVIN STANDARD; PRT; 563 AA.
AC P98091; 1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Submaxillary mucin-like protein.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Submaxillary gland;
RX MEDLINE=90370871; PubMed=2204065;
RA Bhargava A.K., Wotach J.T., Davidson E.A., Bhavanandan V.P.;
RT "Cloning and cDNA sequence of a bovine submaxillary gland mucin-like
RL protein containing two distinct domains."
RT Proc. Natl. Acad. Sci. U.S.A. 87:6798-6802(1990).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC -1- TISSUE SPECIFICITY: TO PORCINE AFOMUCIN.
CC -1- SIMILARITY: Contains 1 WFCC domain.
CC -1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M36192; AAA30657.1; --
DR F01; A36054; A36054.
DR InterPro; IPR006208; Cys_knot.
```

```
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR001007; WFCC.
DR Pfam; PF00007; Cys_knot; 1.
DR SMART; SM00041; CT; 1.
DR SMART; SM00214; VWC; 1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01208; WFCC_1; 1.
DR PROSITE; PS0184; WFCC_2; 1.
KW Repeat; Glycoprotein.
FT DOMAIN 61 158 3 X 11 AA REPEATS.
FT REPEAT 61 71 1.
FT REPEAT 112 122 2.
FT REPEAT 148 158 3.
FT DOMAIN 338 404 WFCC.
FT DOMAIN 471 555 CTCK.
FT DISULFID 471 518 BY SIMILARITY.
FT DISULFID 485 532 BY SIMILARITY.
FT DISULFID 494 548 BY SIMILARITY.
FT DISULFID 498 550 BY SIMILARITY.
FT DISULFID ? 554 BY SIMILARITY.
FT CARBOHYD 28 28 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 394 394 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 477 477 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 563 AA; 58913 MW; AB326CD78ESFFPCF CRC64;

Query Match 5.0%; Score 142; DB 1; Length 563;
Best Local Similarity 20.3%; Pred. No. 0.35;
Matches 75; Conservative 48; Mismatches 169; Indels 78; Gaps 13;

QY 88 EVLHGHPLDNLNRQIGSSVFRVET-----QEDGKHIAVGQNGVETSVL 134
DB 2 KVLQENSPRAHS---GSSHTTEATLLIVSNSTGTLRPEDNTAVAGQATGRVT---53
QY 135 DOEVARLQSIDPEGDKFVFTGGRGGAGHAMVTVASDITEARQRIEELLEPKGTGESKGA 194
DB 54 -----GTTK-VIPGTVAFSSNTSTSLGESRTIGRI-----TGATGT 94
QY 195 GESKGVGELRESNGA-ENTTETQTSSTSLSDPKLMLALGTATGVLGLAATGIVQA 253
DB 95 SKRSSPGS-KTGNATGALSTGTVAPSSNTGATTS-----LCSGETTQGGIKIVTMGVTTG 148
QY 254 LALTPEDSPPTTTPDPAASATETATRDQLTKEAFONPDQKNIDELGNAIPSGVLKDD 313
DB 149 TTIAFGSSNTKATPTFEVTTTETVATET-----TSRHSSDATGSGIQIGI---196
QY 314 VVANIEEQAKAGEEAKQQAENNAQAQKYDEQQAQKQEBLKVSSGAGYGLSGALILGG 373
DB 197 -----TGTSGTTSPPGFGNAETTFKEH-----VTTTETRIILSGTTRGSGTTVIPE 244
QY 374 GIGVAVTAALHRKNQPVQTITTTTTTTTSARTVENKPNANTPAQGNVDTFGSEDIMES 433
DB 245 SSNTGTSTGVCK-----QTSTAVVSGRVTVG-SESSSPGTSKEASETTTGGIGSTTGST 297
QY 434 RRSMASTSS 443
DB 298 SKSNRIITSS 307

RESULT 3
STAU_DROME STANDARD; PRT; 1026 AA.
AC P25159;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Maternal effect protein staufen.
GN STAU.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
```

RN SEQUENCE FROM N.A.
 RX MEDLINE=91300552; PubMed=1712672;
 RA St Johnston D., Beuchle D., Nusslein-Volhard C.;
 RT "Staufen, a gene required to localize maternal RNAs in the Drosophila
 egg.";
 RL Cell 66:51-63 (1991).
 RN [2]
 RP STRUCTURE BY NMR OF 579-646.
 RX MEDLINE=95354674; PubMed=7628456;
 RA Bycroft M., Grunert S., Murzin A.G., Proctor M., St Johnston D.;
 RT "NMR solution structure of a dsRNA binding domain from Drosophila
 RT staufen protein reveals homology to the N-terminal domain of
 RT ribosomal protein S5.";
 RL EMBO J. 14:3563-3571 (1995).
 RN [3]
 RP ERRATUM.
 RA Bycroft M., Grunert S., Murzin A.G., Proctor M., St Johnston D.;
 RL EMBO J. 14:4385-4385 (1995).
 RN [4]
 RP CHARACTERIZATION OF DRBM DOMAINS.
 RX MEDLINE=20183617; PubMed=10716936;
 RA Micklem D.R., Adams J., Grunert S., St Johnston D.;
 RT "Distinct roles of two conserved Staufen domains in oskar mRNA
 RT localization and translation.";
 RL EMBO J. 19:1366-1377 (2000).
 CC -!- FUNCTION: REQUIRED BOTH FOR THE LOCALIZATION OF MATERNAL
 CC DETERMINANTS TO THE POSTERIOR POLE OF THE DROSOPHILA EGG AND
 CC FOR BICOID RNA TO LOCALIZE CORRECTLY TO THE ANTERIOR POLE.
 CC OSKAR PROTEIN IS REQUIRED TO KEEP OSKAR RNA AND STAUFEN PROTEIN AT
 CC THE POSTERIOR POLE.
 CC -!- TISSUE SPECIFICITY: POLAR GRANULES AT THE POSTERIOR POLE OF THE
 CC OOCYTE, AND BY THE TIME THE EGG IS LAID, AT THE ANTERIOR POLE.
 CC -!- DOMAIN: CONTAINS A PROLINE-RICH DOMAIN. THE INSERTION OF THIS
 CC DOMAIN IN THE DRBM 2 DOMAIN IS REQUIRED FOR STAU-OSKAR RNA
 CC LOCALIZATION.
 CC -!- SIMILARITY: CONTAINS 5 DRBM (double-stranded RNA-binding) domains.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M69111; AA73062.1; -.
 DR FIR; A40315; A40315.
 DR PDB; 1STU; 31-JUL-95.
 DR PDB; 1EKZ; 21-AUG-00.
 DR FlyBase; FBgn003520; stau.
 DR GO; GO:0045179; C:apical cortex; IDA.
 DR GO; GO:0003730; F:mRNA 3' UTR binding activity; IDA.
 DR GO; GO:0045034; P:neuroblast cell division; NAS.
 DR GO; GO:0045450; P:pole plasm bicoid mRNA localization; NAS.
 DR GO; GO:0007316; P:pole plasm bicoid mRNA localization; NAS.
 DR InterPro; IPR001159; DS_RBD.
 DR Pfam; PF00035; dsrm; 5.
 DR SMART; SM00358; DSRM; 4.
 DR PROSITE; PS50137; DS_RBD; 5.
 KW Developmental protein; RNA-binding; Repeat; 3D-structure.
 FT DOMAIN 1 311 378
 FT DRBM 1.
 FT DOMAIN 398 557
 FT DRBM 2 (ATYPICAL).
 FT DOMAIN 578 645
 FT DRBM 3.
 FT DOMAIN 711 781
 FT DRBM 4.
 FT DOMAIN 951 1018
 FT DRBM 5.
 FT DOMAIN 3 77
 FT GLN/HIS/PRO-RICH.
 FT DOMAIN 42 46
 FT POLY-GLN.
 FT DOMAIN 47 51
 FT POLY-PRO.
 FT DOMAIN 71 77
 FT POLY-GLN.
 FT DOMAIN 451 455
 FT POLY-PRO.
 FT HELIX 580 590
 FT TURN 591 591

FT STRAND 595 598
 FT STRAND 609 616
 FT TURN 617 618
 FT STRAND 619 627
 FT TURN 628 629
 FT HELIX 630 644
 SQ SEQUENCE 1026 AA; 110347 MW; AESB97624BBF7D0B CRC64;

 Query Match 4.9%; Score 140.5; DB 1; Length 1026;
 Best Local Similarity 20.0%; Pred. No. 0.92;
 Matches 97; Conservative 74; Mismatches 207; Indels 107; Gaps 18;

 QY 2 PIG-NLGHNPVNNISIPAPPLPSQTDGAGRGQLINSTGEL-----GSRALFT----- 49
 DB 501 PLGAHVHGP--NGPPFPSTPPSKITLFGVKQFVGIRTLQQAQKHAARALQVLTKQ 558
 QY 50 -----PVRNSMADSGNDRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGLDNLNRQ 102
 DB 559 AISASEALEDSM--DEGDKXS-----PIQVHEIGIKRNNVHFVKVLRREGPAHMKNFI 611
 QY 103 IGSSVFRVETQEDGKHIAVGORNGVETSVVLSDOEYARLQSIDPEGKDKF-----VFTGGR 158
 DB 612 TACIVGSIIVTEGEGKVKRAAEKMLV-----ELQKLPLTPTKQTPKRIKVKTPGK 667
 QY 159 GGAGHAMVTVAS-----DITEARQRIELLELPKGTGSGSKG 193
 DB 668 SGAAAREGSVVGTDGTQMTGPKPERRKRLNPPKDLIDMDADNPITKLIQLOQTRK--- 724
 QY 194 AGESKGVGELRENSG-----AENTTETQSTSSLSRSDPKLWLAGTATGVLGLAAT 248
 DB 725 --EKEPFELTAKNGNETARRREFVMEVSAGSGTARGTNSKK-LAKRNAQAALFEL--- 778
 QY 249 GIVQALALTPEPDSPTTDPDAAAASATETATRDLTKEAFQNPQNKVINDELGNAIPSG 308
 DB 779 --LEAVQVTP-----TNETQSSECTSATMSAVTAPAVETAEGKVPWATPVGPMGP 830
 QY 309 VL-----KDDVV---ANIE-----FOAKAAGEEAKQOAIENNAQAQKYDE 346
 DB 831 ILILRQNKPAKQDQIVIVKSNVESKEEANKVAEVAEENSNNSNGSSSSGDS 890
 QY 347 QOAKRQBELKVSAGYGLSGALILGGIGGIVAVTAALHRKNQPVQITTTTTTTTISAR 406
 DB 891 QATEAASESALNTSGTSGTSSVSSNVGANTDGNNAHESK--NNTSESSNSTSTNTQSA 948
 QY 407 TVENK 411
 DB 949 GVHMK 953

 RESULT 4
 US45 LACLC
 ID US45 LACLC STANDARD; PRT; 461 AA.
 AC P22865;
 DT 01-AUG-1991 (Rel. 19, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE Secreted 45 kDa protein precursor.
 GN USP45.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 CC NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-37.
 RC STRAIN=MGI363;
 RX MEDLINE=91071599; PubMed=2123812;
 RA van Asseldonk M., Rutten G., Oteman M., Siezen R.J., de Vos W.M.,
 RA Simons G.;
 RT "Cloning of usp45, a gene encoding a secreted protein from
 RT Lactococcus lactis subsp. lactis MGI363.";
 RL Gene 95:155-160 (1990).
 CC -!- SIMILARITY: TO E.PAECIUM P54.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration

Qy	347	QQAQRQEEKLVSSGAGYGLSGALILGGGIGVAVTAALHRKXNPVEQTITTTTTTTTTTT--S	404
Db	406	EPSTSTTTTEVTS-----TSSTVTVTTEPTTTLTSTASTSTTSS	445
Qy	405	ARTVENKPA-----NNTPAQGNVDTPGSEDTMESRBSMASSTSTPFDFSSIG	452
Db	446	TSTVTTSPTSPVTSTVTSSSSSTVTTPSTSTSTSPSSSTVTSTTAPSTSTTG	502
RESULT 6			
Y579	CHLTR		
ID	Y579	CHLTR	STANDARD; PRT; 439 AA.
AC	O84583;		
DT	16-OCT-2001	(Rel. 40, Created)	
DT	16-OCT-2001	(Rel. 40, Last sequence update)	
DT	28-FEB-2003	(Rel. 41, Last annotation update)	
DE	Protein	CT579.	
GN	CT579.		
OS	Chlamydia trachomatis.		
OC	Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.		
OC	NCBI_TaxID=813;		
OX	[1]		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=D/UN-3/Cx;		
RC	MEDLINE=99000809; PubMed=9784136;		
RC	Stephens R.S., Kalman S., Lammel C.J., Fan J., Marache R., Aravind L.,		
RA	Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,		
RA	David R.W.;		
RT	"Genome sequence of an obligate intracellular pathogen of humans:		
RT	Chlamydia trachomatis.";		
RL	Science 282:754-759(1998).		
CC	-!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0808/CT579/TC0868		
CC	FAMILY.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation		
CC	the European Bioinformatics Institute. There are no restrictions on its		
CC	use by non-profit institutions as long as its content is in no way		
CC	modified and this statement is not removed. Usage by and for commercial		
CC	entities requires a license agreement (See http://www.isb-sib.ch/announcement		
CC	or send an email to license@isb-sib.ch).		
CC	-----		
DR	EMBL; AE001328; AAC68181.1; --		
DR	FIR; E71497; E71497.		
DR	PHCI-2DPAGE; O84583; --		
CC	Complete proteome.		
KW	SEQUENCE 439 AA; 44023 MW; F078A41149E08B68		
SO	CR64:		

Query Match	4.7%;	Score 133;	DB 1;	Length 439;
Best Local Similarity	23.8%;	Pred. No. 0.84;		
Matches	73;	Conservative 40;	Mismatches 100;	Indels 94; Gaps 14
Qy	182	LLEPGTGBESKAGESK-----GVGELRESNGAENTTE-----TOTSTS-----TSSLR	226	
Db	53	LQDAQGTGGSKATANTQTSKSSEKAQAASGCTSTTTSVAQASQTATAQAVHGARDSGFN	112	
Qy	227	SPDKLWL--ALCVATGLI-----GLAATGIVQALALTPEPDPSTTTDDPAASATETA	278	
Db	113	SDGSATLPSPPTGEVNGVLKRGMTGLALMLIMTLI-----AAQSAKSWS	158	
Qy	279	TRDOLTKEAFONPDNKVNIDELGNAPISGVLKDDVVANTEEQAAGAEBEAKOQATENNA	338	
Db	159	SSEFQQNQAIQ--NOVMAPEIGNAIRT-----QANHQAATELOAQOSLISG---	204	
Qy	339	QAMKYDEOQAKROBELKVSSGAGYGLSGAILIGGI-----GVAVTAAHLRKNQPVE	391	
Db	205	-----ITNIVGFVS-----VGCGILSASKSLUGLKSAAF-----IN	236	
Qy	392	QTTTTTTTTTTTSARTVE---NKPANNTPAGNVDTPTGESBETWESRRSSMA-----STGS	443	
Db	237	ETASATTTSLSLAKTATYSALDDVAGTATATVGAKTATGAASAASSAATKLTONMMSASK	296	

QY		444	TFFDTSS	450	
			:		
Db		297	TLSTQAS	303	
RESULT 7					
Y572 CHLPN					
ID	Y572	CHLPN	STANDARD;	PRT;	755 AA.
AC	Q9Z7Y1;				
DT	16-OCT-2001	(Rel. 40,	Created)		
DT	16-OCT-2001	(Rel. 40,	Last sequence update)		
DT	16-OCT-2001	(Rel. 40,	Last annotation update)		
DE	Protein CPN0572/CP0177/CPJ0572.				
GN	CPN0572 OR CP0177 OR CPJ0572.				
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).				
OC	Bacteria; Chlamydiae;				
OX	Chlamydiales; Chlamydiaceae;				
CC	Chlamydophila.				
NCBI_TaxID=83558;	[1]	SEQUENCE FROM N.A.			
RN	STRAIN=CWL029;				
RC	MEDLINE=99206606; PubMed=10192388;				
RX	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,				
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.;				
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis";				
RL	Nat. Genet. 21:385-389(1999).				
RP	[2]	SEQUENCE FROM N.A.			
RN	STRAIN=AR39;				
RC	MEDLINE=20150255; PubMed=10684935;				
RX	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,				
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,				
RA	Linhner K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,				
RA	Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,				
RA	Eisen J., Fraser C.M.;				
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia				
RT	pneumoniae AR39.";				
RL	Nucleic Acids Res. 28:1397-1406(2000).				
RN	[3]	SEQUENCE FROM N.A.			
RN	STRAIN=J138;				
RC	MEDLINE=20330349; PubMed=10871362;				
RX	Shibai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,				
RA	Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;				
RA	"Comparison of whole genome sequences of Chlamydia pneumoniae J138				
RT	from Japan and CWL029 from USA.";				
RL	Nucleic Acids Res. 28:2311-2314(2000).				
CC	-1- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0572/CT456/TC0741				
CC	FAMILY.				

```

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; AE001642; AAD18712.1; -;
CC CC EMBL; AE002178; AAF38051.1; -;
CC CC EMBL; AF002547; BAA98778.1; -;
CC CC PIR; B72061; B72061.
CC CC PIR; H86561; H86561.
CC CC PHCI-2DPAGE; Q9Z7Y1; -.
CC CC TIGR; CP0177; -.
CC CC Complete proteome.
CC CC KW SEQUENCE 755 AA; 77578 MW; E57D165BE973AC81 CRC64;
CC CC -----
CC CC Query Match 4.6%; Score 132.5; DB 1; Length 755;
CC CC Best Local Similarity 21.1%; Pred. No. 1.8;
CC CC Matches 104; Conservative 57; Mismatches 166; Indels 167; Gaps 23;
CC CC
CC CC 5 NLGHNPNNVANSIPPAPLPFSQTDAGRGQLINSTGP-----L 42
CC CC ::::: : : : :

```

```

Db 325 DVGLKGINWKTTPADPFSFINETPGGAHSTHTPGCTPGVATVPVNVNVLGGIKVDL 384
Qy 43 GSRAL-----FTPVNSWADSGNRRAS-----DVEPL--- 69
Db 385 GGINLGGITNTVTTEGGGTNTSTKSTSTDDKVSITSTGSGSTIEBTIOFDPPGGED 444
Qy 70 -----PVNPMRLAAAEI-----TLNDGFEVLHDHGPDLTLNRQIGSSVFRVE 111
Db 445 DNAIPGTNTPPPPPPNLSRLLTISNASLNQVLQVNRQH--LNTAYDSNGNSV--SD 500
Qy 112 TOEDGKHIAVGORNGVETSVLSDOEYARLQSIDPEGKDKFVTG--RGGAGHAMVTVAS 170
Db 501 LNQDLGVQVXNSGVNFPVTLPTTG---DTDESGQ---ATGGVTEGGGH----- 546
Qy 171 DITEARORILELLEPKGTGSKGAGE-----SKVGELRESNSGAEN-----T 213
Db 547 -IRNIOR-----NTQSTGQSEGATPTPQPTIAKIVTSLRKANVSSSVLPQPVATTIT 600
Qy 214 TETOT-STSTSLRSDPKLMLALGT-VATGLIGLAATGIVQALALTPEDPSPTTTDPDA 271
Db 601 POARTASTSTSTGIGTSTSTSTGTGTGVSSTQSTGV-----GPTTTTTRSTG 650
Qy 272 ASATTATRDQLTKAFQNPNDQKNVIDELGNAIPSGVLKDDVANIIEQAKAAGEEAKQ 331
Db 651 TSAATTT-----SSASTQTPQ-----APLPSGTRH---VATISLVRNAAG----- 687
Qy 332 QAIENNAQAQKYDQQAQKQBELKV-SSGAGYGLSGALI-----LGGGIGVAVTAALH 384
Db 688 -----RSIVLQCGRSQSFPIPPSGTGTONMGAQLMAAASQVASTLQGVVNQAT 737
Qy 385 RKNQPVQEQTTTTT 398
Db 738 AGSQPSRSSPSPTS 751

RESULT 8
ICEN ERWHE
ID ICEN ERWHE STANDARD; PRT; 1258 AA.
AC P16239;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein.
GN ICEE.
OS Erwinia herbicola.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=549;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M1;
RX MEDLINE=90152370; PubMed=2515997;
RA Warren G.J., Corotto L.V.;
RT "The consensus sequence of ice nucleation proteins from Erwinia
RT herbicola, Pseudomonas fluorescens and Pseudomonas syringae.";
RL Gene 85-239-242(1989).
CC -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC -!- CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS 126 IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC -!- A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/)

```

```

CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; M26382; AAA24823.1; -.
DR PIR; JQ0188; JQ0188.
DR HSP; P06620; IINA.
DR InterPro; IPR000258; Ice_nucleatn.
DR Pfam; PF00818; Ice_nucleation; 65.
DR PRINTS; PRO0327; ICENUCLEATN.
DR PROSITE; PS00314; ICE_NUCLEATION; 45.
KW Ice nucleation; Repeat; Outer membrane.
FT DOMAIN 162 1217 OCTAPEPTIDE PERIODICITY.
SQ SEQUENCE 1258 AA; 125084 MW; 590E8A130077FBD4 CRC64;

Query Match 4.6%; Score 132.5; DB 1; Length 1258;
Best Local Similarity 20.4%; Pred. No. 3.4;
Matches 110; Conservative 62; Mismatches 226; Indels 141; Gaps 20;

Qy 26 TDGAGRGQLI---NSTGPLGSRALFTPVNSWADSGNRRASDVPGIIVNPMRLAAAEI- 81
Db 183 TETAGDSSTLIAGYGTGTAGADSTLVAGYGSTQTAGE-ESSQWAGYGSTQGMKGSDLT 241
Qy 82 -----TLNDGFEVLHDHGPDLT-----LNRIQSSVFRVETOEDGKHIAVGORN- 125
Db 242 AGYGTGTAGDSSSLIAGYGTGTAGDSSSLIAGYGT-----QTAQKGSDLTGYGTGT 297
Qy 126 -GVETSVVLSQDYARLQSIDPE-----GDKKFVFTGGRGAG----- 162
Db 298 AGADSSLIAG---YGSTQTAGEESTQTAGYGTGTQAKGSDLTAGYGTGTAGDSSSLIA 354
Qy 163 -----HAMVTASDITEARORILELLEPKGTGSKGAGESKVGELRESNSGAEN 212
Db 355 GYGTGTAGDSSSLIAGYGTGTQAKGSDLTAGYGTGTAGDSSSLIAGYGTGTAGEES 414
Qy 213 T-TETQSTSTSSLSRSDPKLMLALGTAVTGLIG-----LAATGIVQALALTPEDSPPT- 265
Db 415 TQTAGYGTQTAQKGS-----LTAGYGTGTAGDSSSLIAGYGTQ-----TAGDSSSLTA 466
Qy 266 -----TDPDAASATETATRDQLTKAEF-----RKNQPVQEQTTTTTTT 295
Db 467 GYGTGTQTAQKGSDLTAGYGTSTAGYESSLIAGYGTGTAGYGTSTLTAGYGTQTAQNES 526
Qy 296 VNIDELGNAPSGVLKDDVANIIEQAKAAGEEAKQAIENNAQAQKYDQQAQKQBEL 355
Db 527 DLITGYGTSTAGA-----NSSLIAGYGTGTATASVNSVLTAGYGTGTAREGSDL 576
Qy 356 KVSSGAGYGLSGAL-----ILGGGIGVAVTAALH-----RKNQPVQEQTTTTTTT 399
Db 577 T-----AGYGTGTAGDSSSLIAGYGTGTATASVNSVLTAGYGTGTAREQSVLTGYGSTS 632
Qy 400 TTTTSARTVENKPNANTPAQGNVDTPGSEDTMESRRSS-----MASTSTFFDTSSIGG 453
Db 633 TAGADSSSLIAGYGTGTATAGYNSILTAGYGTGTQTAQEGSDLTAGYGTSTAGDSSSLIAG 691

RESULT 9
YEEJ ECO57
ID YEEJ ECO57 STANDARD; PRT; 2660 AA.
AC Q8X8V7; Q8X2B9; Q8X2C0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Hypothetical protein yeej.
GN Z3135 OR ECS2775/ECS2776.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glaesner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

```

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamocousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7";
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIND 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Iehii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasaki C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic *Escherichia coli*
RT O157:H7 and genomic comparison with a laboratory strain K-12";
RL DNA Res. 8:11-22 (2001).
CC -!- SIMILARITY: Contains 16 Big-1 domains.
CC -!- SIMILARITY: BELONGS TO THE INTIMIN/INVASIN FAMILY.
CC -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC frameshift in position 1315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE005423; BAG57041.1; -;
DR EMBL; AP002559; BAB36198.1; ALT_FRAME.
DR EMBL; AP002559; BAB36199.1; ALT_FRAME.
DR InterPro; IPR003344; Big 1.
DR InterPro; IPR003535; Intimin.
DR InterPro; IPR000601; PKD_domain.
DR Pfam; PF02369; Big_1; 16.
DR PRINTS; PR01369; INTIMIN.
DR SMART; SM00634; BID 1; 16.
DR SMART; SM00089; PKD 8.
KW Hypothetical protein; Repeat; Complete proteome.
FT DOMAIN 738 834
FT DOMAIN 840 929
FT DOMAIN 931 1033
FT DOMAIN 1042 1132
FT DOMAIN 1134 1236
FT DOMAIN 1245 1335
FT DOMAIN 1337 1439
FT DOMAIN 1448 1539
FT DOMAIN 1548 1652
FT DOMAIN 1653 1750
FT DOMAIN 1751 1855
FT DOMAIN 1856 1957
FT DOMAIN 1963 2056
FT DOMAIN 2065 2156
FT DOMAIN 2157 2252
FT DOMAIN 2254 2355
FT DOMAIN 2355 2455
SQ SEQUENCE 2660 AA; 280062 MW; 01EB92A08F5C09D2 CRC64;
Query Match 4.6%; Score 131; DB 1; Length 2660;
Best Local Similarity 19.7%; Pred. No. 11;
Matches 123; Conservative 88; Mismatches 224; Indels 190; Gaps 25;
QY 33 GQLINSTGLPSRALFTFVRNSMADS-----GDNRASDVP 68
DB 1176 GITVTFMPQDVAAFTLENNGIAITQANGBAHVTLKGKAGTHVTATLGNNSAQP 1235
QY 69 L-----PYNRLAASEI-----TLNDGF-----EVLHDHGLDTLNR 101
DB 1236 VTFVADKSDAVVVLQTSKAEILNGVDVETTLTATVKDFDNAVXDLQVTFSTNADT--- 1292
QY 102 QIGSVFRVTEQDQ--KHIAVGORGNGVETSVLSDQBYARLQSIDPEGDKDFVFTGGRG 159

Db 1293 QLSOS--KSNVNDSGVAEVTFGVTLGVHTA-----EATLPNGNNDTKIVNIAP 1339
QY 160 GAGHAMVT-----VASDITEARQRIELELPKG-----TGESK 192
Db 1340 DASNAQVTLNIPAQOVVTTNNSDSVQLTATVKDPNSHPVAGITVNTFTMPQDVAANFTLENN 1399
QY 193 GAGESKGVGELRESGAENTTETQTSSTSLSD--PKLWLALGTVATGLGLAATGI 250
Db 1400 GIATTOANGEAHVTLKGKAGTHVTATLGNNSNTSDSQPVTFVADKTSALVVLQISKNEI 1459
QY 251 VQALALTPEDSPPTTDPDAASATETATRDQLTKEAFQNPDPDNOKVNDLGNALPSGV 310
Db 1460 T-----GNGVDSATLTAT-----VKDQF-----DNEVNNLPVTFSTASSGLT 1496
QY 311 KDDVIANIE-----QAKAA-----GEEAKQQAIIENNAQAQKY-----DEQAKRQEELKV 357
Db 1497 LTPGESNTNESGIAQATLAGVAFGEQVFTASLANNGASDNKTVHFIGDTAAAKIELTPV 1556
QY 358 SSGAGYGLSGALILG---GGIGVAVTAALHRKNO-PVEQTTT-----TTTTTSARTVENKPA 413
Db 1557 PD-----SIIAGTPQNSGSGVITATVVDNNGFPVKGVTVNFTSNAATAEMTNGQAV 1608
QY 414 NN-----TPAQGNVDTPGSEDTMESRRSSMASTSTFFDTSIGGCPRIEMLMLKH 464
Db 1609 TNEQKGATVYTNTRSSIEGARDPTVEASLENGSSTLSTSVNADASTAHLTLL----- 1664
QY 465 RCMIRRCRLILRLPRWIGIQISVVYSTIQHPDRDITDNGARLLGNPSAGIQSTYARLA 524
Db 1665 -----QALFDTVS--AGDTNLYIEVKDNGVNGVPOQEVTL 1699
QY 525 LSGGLRHDMGGLTGGSNSAVNTSN 549
Db 1700 VS-----PSEGVTP-SNNAYITTNH 1718
RESULT 10
ICEN PANAN
ID - ICEN PANAN STANDARD; PRT; 1034 AA.
AC Q47879;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ice nucleation protein inaU.
GN INAU.
OS Pantoea ananas (Erwinia uredovora).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pantoea.
OX NCBI_TaxID=553;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KUIN-3;
RX MEDLINE=94264407; PubMed=7764866;
RA Michigami Y., Watabe S., Abe K., Obata H., Arai S.;
RT "Cloning and sequencing of an ice nucleation active gene of *Erwinia*
RT uredovora.";
RL Biosci. Biotechnol. Biochem. 58:762-764 (1994).
CC -!- FUNCTION: ICE NUCLEATION PROTEINS ENABLE BACTERIA TO NUCLEATE
CC CRYSTALLIZATION IN SUPERCOOLED WATER.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- DOMAIN: CONTAINS IMPERFECT REPEATS OF A CONSENSUS OCTAPEPTIDE
CC A-G-Y-G-S-T-X-T; FURTHER ON A 16-RESIDUE AND A REGIONAL 48-RESIDUE
CC PERIODICITY IS SUPERIMPOSED.
CC -!- MISCELLANEOUS: A STRUCTURAL MODEL IS SUGGESTED IN WHICH THE ICE
CC NUCLEATION PROTEIN DISPLAYS A SYMMETRY RELATED TO THAT OF ICE.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL ICE NUCLEATION PROTEIN
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M95192; AAA28607.1; ALT_INIT.
 DR EMBL; X67239; CAA47664.1; --
 DR EMBL; AE003731; AAF55790.1; --
 DR EMBL; AE003731; AAF55791.1; --
 DR PIR; A44067; A44067.
 DR FlyBase; FBgn0001169; H.
 DR GO; GO:0003714; F:transcription co-repressor activity; IDA.
 DR GO; GO:0007219; P:N receptor signaling pathway; NAS.
 DR GO; GO:0008052; P:sensory organ determination; IMP.
 KW Developmental protein; Nuclear protein; DNA-binding;
 KW Alternative splicing.
 FT DOMAIN 115 123 THR-RICH.
 FT DOMAIN 642 648 POLY-SER.
 FT DOMAIN 879 891 POLY-ALA.
 FT DOMAIN 937 946 POLY-ALA.
 FT DOMAIN 964 974 ALA-RICH.
 FT DOMAIN 979 1008 HIS/PRO-RICH (PRD MOTIF).
 FT VARSPLIC 1 18 Missing (in isoform 2).
 FT CONFLICT 151 151 S -> A (IN REF. 2).
 FT CONFLICT 680 680 S -> F (IN REF. 3).
 FT CONFLICT 702 703 QH -> LL (IN REF. 2).
 FT CONFLICT 891 891 A -> R (IN REF. 2).
 FT CONFLICT 964 967 AAVA -> RLLP (IN REF. 2).
 FT CONFLICT 974 974 MISSING (IN REF. 2).
 SQ SEQUENCE 1077 AA; 111039 MW; A94BF1A27579E2F1 CRC64;

Query Match 4.6%; Score 130; DB 1; Length 1077;

Best Local Similarity 18.9%; Pred. No. 3.9;

Matches 145; Conservative 66; Mismatches 244; Indels 312; Gaps 29;

Qy 7 GHPNVNNSIPPAPP-----LPSOTDGGGGRGOLINSTGPLGS--RALF 48
 Db GLNGSSSTPPLPPPLPALSRITPTTTPSSSSSTASNGFLPRAKTPKSSINAAAS 151
 Qy 49 TPVRNSMADSGDNRAS-DVPG-----LPVNPMLAASEIT-----LNDGF 87
 Db 152 AAVAASVVGATASKPTIDVLGGVLDYSLGGAATGSLPTTAVVAAAAGATAKIGKNSGG 211
 Qy 88 EVLHDGHPDLTL-NRQIGSSVFRVETQEDKHIAVGQRNGVETSVVLSDQRYARLQSIDP 146
 Db 212 SFDMGRTPISTHGNNSGGYGRQLQFPKDGKFI-----ELARSKDGDK 255
 Qy 147 EGKDKFVFTGGR--GGAGHAMVTVASDITEARQRIELLEPKGTGSKGAGESKGVGELR 204
 Db 256 SGWVSVTRKTRPPSAATSATVTPTSAVTTA-----YPRN-----290
 Qy 205 ESNSGAENTETQTSTSTSLRSDP-----KLWLAGTVAATGLIGLA 246
 Db 291 -----ENSTLSFSDNNSIQSPWQDQPKQSRPRRGISKLSLFFHPRNSTLGRA 344
 Qy 247 ATGIVQALALTPEPDSPTTDDP-----AAASATE---TATRDQL 283
 Db 345 A--LRTAARKRRRPHPELTTSDDQPIFAVAIKAEAGDDTLKAEAAEAVETENVAVDTT 402
 Qy 284 TKE-AFQNPD-----LGNAPSVGLKDDVVANTEEQAAGAEAKQAQATENNAQAOKKYDEQ 297
 Db 403 TNEIKIEKPTIKGEDDAERLEKPKKAVSDSESKEASPGQVPEQPKDETVDVEMKM 462
 Qy 298 IDE-----LGNAPSVGLKDDVVANTEEQAAGAEAKQAQATENNAQAOKKYDEQ 347
 Db 463 TSEDEEPMTELPRITNAV-NGDNGDLKASLGKPKPKPKAKULSSI-----IQKLDSV 516
 Qy 348 QAKRQEELKSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVQTTT-----395
 Db 517 PARLEQMSKTSAVIASITTTSSDRIGGGLSHALT---HKVSPSSATAAGRLVEVHTQHS 573

Qy 396 -----TTTTTTTSARTVENKPNNTPAQ 419
 Db 574 PRKRILLREPEKVSLEDNGCVNNGSGGAGGKRSRAKGTSTSSPAGKASPMNLAPPQ 633
 Qy 420 GNVD-TPGSE-----DTMESRRSSM-----ASTSSTFFDTSSIGGPPCRIRMLM 461
 Db 634 GKPSPPSGSSSSSTPATLSTQPTRLNSSYSIHSLGLGGSGSSSSSSSGKKC-----687
 Qy 462 LKHRCMIRRCRLILRLPRIGIQISVYVSTIQHPPTDNGCARLLGNPSAGIQSTYA 521
 Db 688 -----GDHPAAIISNVHHPQHS-----MYDPSS---SSYP 714
 Qy 522 RLALSGSLRHDGMLTGG-----SNSAVNTSNPPAPGS-----HRF 558
 Db 715 RALLTSPKSPDVSGSGGKGKSPSHTGKTKRSPYSAGSPVDYGHSF 761
 RESULT 12
 DAB_DROME STANDARD; PRT; 2411 AA.
 AC P98081;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Disabled protein.
 GN DAB.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
 RX MEDLINE=33194063; PubMed=7680635;
 RA Gertler F.B., Hill K.K., Clark M.J., Hoffmann F.M.;
 RT "Doseage-sensitive modifiers of Drosophila abl tyrosine kinase
 RT function: prospero, a regulator of axonal outgrowth, and disabled, a
 RT novel tyrosine kinase substrate."; Genes Dev. 7:441-453(1993).
 RL Genes Dev. 7:441-453(1993).
 CC -|- FUNCTION: TOGETHER WITH ABL INVOLVED IN EMBRYONIC NEURAL
 CC DEVELOPMENT.
 CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=P98081-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=P98081-2; Sequence=VSP_004185;
 CC -|- TISSUE SPECIFICITY: CNS AXONS AND BODY WALL MUSCLES.
 CC -|- DEVELOPMENTAL STAGE: EMBRYONIC AXONOGENESIS.
 CC -|- PTM: PROBABLY PHOSPHORYLATED BY THE ABL TYROSINE KINASE.
 CC -|- SIMILARITY: Contains 1 PID domain.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

EMBL; L08845; AAB08527.1; ALT_SEQ.

PIR; A46299; A46299.

FlyBase; FBgn0000414; Dab.

GO; GO:0005737; Cytoplasm; IDA.

InterPro; IPR006020; PTB_PID.

Pfam; PF00640; PID; 1

SMART; SM00462; PTB; 1.

PROSITE; PS01179; PID; 1.

Alternative splicing; Phosphorylation; Repeat.

DOMAIN 46 196

DOMAIN 1689 1801 REPEAT-RICH REGION.

```

DE DE Flagellar hook-associated protein 2 (HAP2) (Filament cap protein)
DE DE (Flagellar cap protein).
DE GN FLID OR VC2140.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=El Tor N16961 / Serotype O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Hatt D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Emolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
CC -I- FUNCTION: REQUIRED FOR THE MORPHOGENESIS AND FOR THE ELONGATION OF
CC THE FLAGELLAR FILAMENT BY FACILITATING POLYMERIZATION OF THE
CC FLAGELLIN MONOMERS AT THE TIP OF GROWING FILAMENT. FORMS A CARPING
CC STRUCTURE, WHICH PREVENTS FLAGELLIN SUBUNITS (TRANSPORTED THROUGH
CC THE CENTRAL CHANNEL OF THE FLAGELLUM) FROM LEAKING OUT WITHOUT
CC POLYMERIZATION AT THE DISTAL END (BY SIMILARITY).
CC -I- SUBUNIT: Homopentamer (By similarity).
CC -I- SUBCELLULAR LOCATION: Flagellar.
CC -I- SIMILARITY: BELONGS TO THE FLID FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE004287; AAF95285.1; -
DR FIC: G82111; G82111.
DR TIGR: VC2140; -.
DR InterPro: IPR003481; Flid.
DR Pfam: PF02465; Flid; 1.
DR Flagella; Coiled coil; Complete proteome.
FT DOMAIN 354 419 COILED COIL (POTENTIAL).
SQ SEQUENCE 666 AA; 72330 MW; 6794EFC34A2A0D2 CRC64;

Query Match 4.5%; Score 129; DB 1; Length 666;
Best Local Similarity 19.6%; Pred. No. 2.4;
Matches 107; Conservative 77; Mismatches 221; Indels 140; Gaps 24;

Qy 1 MPIGNLGHNP -NVN---NSIPAPPPLPSQTDGAGRGQLINSGPLGS-RALETPVRNS 54
Db 1 MSLGPMGMNVTGFDINGVMVSKVSAERPVKQQRIDNERTNIDTISAYGRURLESIDTMKNL 60
Qy 55 MADSGDNRAVDVPLGPNMRLAAASEITLN-----DGFEVLHDI----- 93
Db 61 MTQPRQSKAFARVKDTISNQVVSATATTEAIGNYSVDVLQLAQSHKIASEVLDKDAKF 120
Qy 94 GPLDITLNRQIGSSVFRVETQEDGKHIAV-----GQRN--GVETSVLVSDQYARL-OSID 145
Db 121 GP-GLKHLISLGDKSFTLDVQGNKSLVDIVRGINGEKSNPGVRASII-NDVGGPRLIVASN 178
Qy 146 PEGKDKVFVFTGGRGAGHAMVTV-----ASDITEARQILLELPKGTGESK----- 192
Db 179 VSGKDHVSQMSAQAEPPNPKLEKYKTLQQRVRLDKARAQOQLIAPLTPEQOKVAAKV 238
Qy 193 -----GAGESKGVGEELRESNGAENTTTQTSTSTSS 224
Db 239 AEKIGDARLVQDVEAQSIRSAQSAQGAAGEALNAGELTETES---AVKAAANAASEAKKY 295
Qy 225 LRSDPKL--WLALGTVAITGLIGLAATGIVQALAITPEPDSFTTTDPDAA-----ASATE 276

```

	179	VSGKDSHVNSAQAEPCGNPLKQLEYKTLEQRVRLERAKAAQQILIAPLTPEQQOKVAKV	233
D _b			
	193	-----GAGESKGVGELRESNGSAENTTTQTSTSS	224
Q _y			
	239	AEKTGDAARLVDEQAQEIRSAASQAGAAEALNAGELITES---AVKAAANAAASEAKKY	295
D _b			
	225	LRSDFKL--WIALGTATGLIGLGAATGIQVALALTPEPDPSTTTDPA	276
Q _y			

Db 296 IRPDRIPGWT-----TASGTLDSYWEPEELDAQQKKAADVPGWSNTASG 344
Qy 277 TATRDQLTKEAFQNPQKNVID--ELGNAIPSGVLKDDVVANIEBQAKAAGEBAKQAI 334
Db 345 TLLDSYVTPQAKLEKQAKAEAAIRSG-----KMTPEBAKQAR 390
Qy 335 ENNAQAQKDYEQAKRQBEELKVSQA--GYG-----LSGALILGGGIGVAVTAALH 384
Db 391 AKLSPEERAYIEQVEKAQAALNAQAQSAFDGCGMTEVQSAQDSMVVLDG-----VATLS 444
Qy 385 RKNQPVQEQTTTTTTTTTTSARTVENKPNATPAQGNVDTPGSDTWSRSMASST 444
Db 445 SNNNIE--NAIEGVNLTGKTDNRQP-----PAIEIGIEY--DRERVNRDIEQVVAAYNQ 496
Qy 445 FFDTS 449
Db 497 FFGTS 501

RESULT 14

SP4_HUMAN
ID SP4_HUMAN STANDARD; PRT; 784 AA.
AC Q02446; Os0402; PRT; 784 AA.
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor Sp4 (SPR-1).
GN SP4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=93087156; PubMed=1454515;
RA Hagen G., Mueller S., Beato M., Suske G.;
RT "Cloning by recognition site screening of two novel GT box binding
RT proteins: a family of Spl related genes";
RL Nucleic Acids Res. 20:5519-5525(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Ozersky P., Holmes A.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Binds to GT and GC boxes promoters elements. Probable
CC -!- transcriptional activator.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: ABUNDANT IN BRAIN.
CC -!- SIMILARITY: BELONGS TO THE SP1 FAMILY OF C2H2-TYPE ZINC-FINGER
CC PROTEINS.
CC -!- SIMILARITY: Contains 3 C2H2-type zinc fingers.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X68561; CAA48563.1; -.
CC EMBL; AC004595; AAD12226.1; -.
CC PIR; S26638; S26638.
CC HSP; P08047; ISPI.
CC TRANSFAC; T02339; -.
CC Genew; HGNC:11209; SP4.
CC MIM; 600540; -.
CC GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; TAS.
CC GO; GO:0003713; P:transcription co-activator activity; TAS.
CC GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; TAS.
CC InterPro; IPR007087; Znf_C2H2.
CC Pfam; PF00096; zf-C2H2; 3.

ProDom; PD000003; Znf_C2H2; 2.
SMART; SM00355; Znf_C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2_2; 3.
KW Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW DNA-binding; Nuclear protein; Repeat.
FT ZN_FING 647 671 C2H2-TYPE 1.
FT ZN_FING 677 701 C2H2-TYPE 2.
FT ZN_FING 707 729 C2H2-TYPE 3.
FT DOMAIN 7 11 POLY-GLU.
FT DOMAIN 12 19 POLY-ALA.
FT DOMAIN 122 130 POLY-SER.
FT DOMAIN 185 188 POLY-SER.
FT CONFLICT 197 197 K -> Q (IN REF. 2).
FT CONFLICT 379 380 HA -> QP (IN REF. 2).
FT CONFLICT 386 386 Q -> A (IN REF. 2).
SQ SEQUENCE 784 AA; 82025 MW; 3C4EAE28CB2B81FB CRC64;
Query Match 4.5%; Score 129; DB 1; Length 784;
Best Local Similarity 20.0%; Pred. No. 3;
Matches 112; Conservative 71; Mismatches 234; Indels 144; Gaps 21;
Qy 3 IGNLGHNPVNNNS---IPPAPPL-----PSQ-----TDGAGGRGOLINSTGPIG-SRAL 47
Db 58 IGTGEGNATGQQQIIIDPSGLVQLNQPOQLBELVTTQLAGNAWLAVSTPPASKENVV 117
Qy 48 FTPVRNSMADSGDNRAVDPLPVPNMLAASETLNDGFVLDHGHGFLDTLNRQIGSSV 107
Db 118 SQPASSSSSSSSNNNGS-----ASPTKSGNSSTPGQFQVIOVQNPSSGVQVIVPQL 171
Qy 108 FRVETQDGHKHAIVGQNGVETSVLSQDQYVARLOSIDPEGKDFVFT-GRGGAGHAMV 166
Db 172 QTVEGQOI-----QINPTSSSLQDLQGGQIKLIS---AGNNQAILTAANRTASGNILA 221
Qy 167 -----TVASDITEARQILLELPKGT-----GSK 192
Db 222 QNLANTVPVQIRPGVSIPIQLQTLPGTAQVVTLPINIGVTLALPVINNVAGGGTG 281
Qy 193 GAGSKGVGLRESNGAENTTQTSTSSLSRDPKLMALGTAVT-----GLIGL 245
Db 282 QVGPAATADSGTNGNLVSTPTNTTTSATMPESSSTTCTTASTSTSTSDTLVSS 341
Qy 246 AATGIVQALALTPEDSPPTTDPDAASATETATR--DQLTKEAFQNPQKNVDEIG- 302
Db 342 ADTG---QYASTSASSERTIEESQTPAATESEASQSSQLHANGWQNDQNSLSLQQVQI 398
Qy 303 -----NAP-----SG-----VLKDDVANIIEQA----- 322
Db 399 VGQPILOQIQPQQQIIQAIPQSPQSQSGGTITQIQPQPLQNVQLQAVNPTQVLIRA 458
Qy 323 ---KAAGEEAKQAIENNAQAOKKYDQQAOKRELVSSGAGYGLSGALILGGGIGVAV 379
Db 459 PTLTPSGQISWQTVQVQNIQSLNQLQVQNAQLSQQLITPVSS---SGTTLAQIAPVAV 515
Qy 380 TAALHRKNQPVQEQTTTTTTTTTTSARTVENKPNATPAQGNVDTPGS-----EDTME 432
Db 516 AGA-----PITLNTAQLASVPNLQTVSVANLGAAGVQVQVPTTITSVAGQQQDQGVK 569
Qy 433 SRRSSMASTSTPFDTSISGG 453
Db 570 VQQTATAPV-----TVAVGG 584
RESULT 15
PMPC_CHLMU
ID PMPC_CHLMU STANDARD; PRT; 1460 AA.
AC Q9EJY1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable outer membrane protein pmc precursor (Polymorphic membrane
DE protein C).
GN PMPC OR TC0695.

OS Chlamydia muridarum.
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83560;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MoPn / Ni99;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39";
RL Nucleic Acids Res. 28:1397-1406(2000).
CC -!- SUBCELLULAR LOCATION: CELL WALL SURFACE (ELEMENTARY BODIES)
CC (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE PMP OUTER MEMBRANE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE002338; AAF39511.1; -;
DR PIR; D81675; D81675.
DR TIGR; TC0695; -;
DR InterPro; IPR006315; Autotransport.
DR InterPro; IPR003368; Chlamydia_PMP.
DR Pfam; PF02415; DUF145; 2.
DR TIGRFAMs; TIGR01414; autotrans_bar1; 1.
DR TIGRFAMs; TIGR01376; POMP_repeat; 8.
KW Outer membrane; Signal; Multigene family; Complete proteome.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 1460 PROBABLE OUTER MEMBRANE PROTEIN PMP.
FT SEQUENCE 1460 AA; 154781 MW; C1F033E7907AC3AC CRC64;
SQ

Query Match 4.5%; Score 129; DB 1; Length 1460;
Best Local Similarity 20.5%; Pred. No. 6.6;
Matches 99; Conservative 58; Mismatches 191; Indels 136; Gaps 21;

QY 13 NNSIIPA--PPLPQTGAGR--GQLNSTGLGSRALPTPVNSMADSGDNR--ASD 65
DB 276 NTVTPPAANKPELPDPGSGNGDSDSGNTDSNE-SNPNNASNNTGSENELSS 334

QY 66 VPG--LPVNPRLAAEITLNDGFVLHDHGLDPLNRLQIGSSVFRVETQEDGKHIAVGQ 123
DB 335 TPSAQLP-NPATPFLSSVSTNS-----QPIDT-----EPENAWHAESGGGAIYSK GK 381

QY 124 RNVETSVVLSQDEYARLQSIDPEKDPVF-----TCGRGAGHAMVTVASD 171
DB 382 LSIASSKEVVFDPHNGATKNGGAIFGEEIEALEKIALKFLDSNTTGERGAIHAKTVTLSD 441

QY 172 ITEARQRILELEPKGTGSKGKSGVGLRESGAEENTTQTSTSSLRSDPKL 231
DB 442 IKNT-----LIFVNTAKTPENSLKS-SQLNNQNPSEEHQDTSEGEESQLETSP-- 492

QY 232 WLALGTVATGLIGLAATGIVQALALTPEPDSPTTTPDAAA-----ATETATRDQLTKE 286
DB 493 -----ITNQDSASSHVAIFRSIAASSSQSNSE 519

QY 287 AFQNPDKNTDELGNAT-PSGVLKDDVVANIEQAKAAGEAQQAIEENNAQAKKYD 345
DB 520 NIPNADGSTSAGDAGSSQSPGSDSSINHVIGGAIYGEAVK---IENL----- 568

QY 346 EQQAKQEELKYSVGAGY-----LSGAL--ILGGGIGVAVTAALHKKQNPVEQ 392
DB 569 -----SGYTFNNNAVDHQISGSTDVLGGAIYAKTSLTIDSGN----- 608

QY 393 TTTTITTTTTSARTVENKPNNTPAQGNVDPFGSEDTEW---SRRSSMASTSTFTDTS 449

Db 609 -----SSGTTTFSENTTSSKSTTCQVAGGAIFSPSVTITTPVTFSKNSAINATTSSKKT- 663
QY 450 SIGG 453
Db 664 -FGG 666

Search completed: December 16, 2003, 08:53:53
Job time : 24.694 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 26.7392 Seconds
(without alignments)
2010.470 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPIGNLGHNPVNNNSIPPAP.....SNSAVNTSNNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2624.5	92.1	558	2 A98199	translocated intim
2	2624.5	92.1	558	2 E86045	probable transloca
3	170	6.0	1229	2 T25697	hypothetical prote
4	165.5	5.8	2232	2 T34434	hypothetical prote
5	162.5	5.7	1291	2 T13389	hypothetical prote
6	152.5	5.3	971	2 B90835	probable tail fibe
7	152.5	5.3	973	2 C85693	probable membrane
8	148.5	5.2	770	2 T51024	related to C2H2 zi
9	148.5	5.2	1246	2 G89287	protein H39E23.1 [
10	148	5.2	1192	2 T18611	probable serine/th
11	146.5	5.1	1122	2 G64887	probable tail fibe
12	144.5	5.1	3507	2 T34513	hypothetical prote
13	143.5	5.0	1275	2 T33369	hypothetical prote
14	143.5	5.0	1630	2 A33577	ascites sialoglyco
15	142.5	5.0	3013	2 AB0480	probable invasiv
16	142	5.0	563	2 A32634	mucin homolog - bo
17	141.5	5.0	1962	2 A32634	lactocepin (EC 3.4
18	140.5	4.9	1026	1 A40315	maternal effect pr
19	140.5	4.9	1829	2 T24583	hypothetical prote
20	139	4.9	2468	2 A83412	hypothetical prote
21	137	4.8	837	2 T02761	outer arm dynein i
22	137	4.8	4558	2 C82199	RTX toxin RtxA VC1
23	136	4.8	949	2 D90803	Aida-I adhesin-lik
24	136	4.8	1005	2 H85611	probable adhesin Z
25	134	4.7	2514	2 F81045	hemagglutinin/hemo
26	133.5	4.7	796	2 T21460	hypothetical prote
27	133	4.7	439	2 E71497	hypothetical prote
28	133	4.7	1589	2 T42233	submaxillary mucin
29	132.5	4.6	461	2 JN0097	secreted 45K prote

ALIGNMENTS

RESULT 1

A98199
translocated intimin receptor Tir [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: A98199
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: A98199
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-558 <HAY>
A:Cross-references: PIDN:BA000007; PIDN:BA037984.1; PID:gl3364036; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC84561

Query Match		92.1%	Score 2624.5;	DB 2;	Length 558;
Best Local Similarity		94.0%	Pred. No. 2.2e-148;		
Matches 528;		Conservative 3;	Mismatches 24;	Indels 7;	Gaps 3;
Qy	1	MPIGNLGHNPVNNNSIPPAPLP	PSQTCAGRGQLINSTGPLGSRALFPPVNSMADSGD	60	
Db	1	MPIGNLGHNPVNNNSIPPAPLP	PSQTCAGRGQLINSTGPLGSRALFPPVNSMADSGD	60	
Qy	61	NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPDLTINRQIGSSVFRVETQEDGKHIA	120		
Db	61	NRASDVPGLPVNPMRLAASEITLNDGFEVLHDHGPDLTINRQIGSSVFRVETQEDGKHIA	120		
Qy	121	VGQRNGVETSVLSDQYVARLQSIDPEKDKFVFTGGRGGAGHAMVTVASDITEARQRI	180		
Db	121	VGQRNGVETSVLSDQYVARLQSIDPEKDKFVFTGGRGGAGHAMVTVASDITEARQRI	180		
Qy	181	ELLSPKGTGSGKAGESKGVCELRESNGAENTTQTSTSTSLRSPKWLALGTAT	240		
Db	181	ELLSPKGTGSGKAGESKGVCELRESNGAENTTQTSTSTSLRSPKWLALGTAT	240		
Qy	241	GLIGLAATGIVQALALTPEDSPITTTDPDAASATATATRDQLTKEAFQNPNDQKVNIDE	300		
Db	241	GLIGLAATGIVQALALTPEDSPITTTDPDAASATATATRDQLTKEAFQNPNDQKVNIDE	300		
Qy	301	LGNAPISGVLKDDVVANIEEQAKAAGEAKQQAENNAQAQKYDEQAKQEELKVSSG	360		
Db	301	LGNAPISGVLKDDVVANIEEQAKAAGEAKQQAENNAQAQKYDEQAKQEELKVSSG	360		
Qy	361	AGYGLSGALILGGGIGVAVTAALHRKNQPVETTTTTTTTTTTTTTARTVENKPNNTPAQG	420		
Db	361	AGYGLSGALILGGGIGVAVTAALHRKNQPVETTTTTTTTTTTTTTARTVENKPNNTPAQG	420		

```
QY 421 NVDTFGSEDTMESRRSSMASTSTFFDTSIG---GPCRIRMLMLKHKRCMIRRCRLILI 477
DB 421 NVDTFGSEDTMESRRSSMASTSTFFDTSIGTGVQNPADVKTSL-HDSQVPTNSNTSV 479
QY 478 RLFRWIGIQISVYVYTIQHPRTDNDGARLLGNPSAGIQSTYARLALSGGLRHDHMGGLT 537
DB 480 QNM---GNTDSVYVYTIQHPRTDNDGARLLGNPSAGIQSTYARLALSGGLRHDHMGGLT 536
QY 538 GGSNAVNTSNNPPAPGSHRFV 559
DB 537 GGSNAVNTSNNPPAPGSHRFV 558

RESULT 2
E86045
probable translated intimin receptor protein tir [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E86045
R:Perma, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86045
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-558 <STO>
A:Cross-references: GB:AE005174; NID:g12518449; PIDN:AAG58825.1; GSPDB:GN00145; UMG:Z51
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: tir

Query Match 92.1%; Score 2624.5; DB 2; Length 558;
Best Local Similarity 94.0%; Pred. No. 2.2e-148;
Matches 528; Conservative 3; Mismatches 24; Indels 7; Gaps 3;

QY 1 MPIGNLGNPNVNNISPPAPLPSTQDAGGRGQLINSTGLPSRALFTPVNSMADSGD 60
DB 1 MPIGNLGNPNVNNISPPAPLPSTQDAGGRGQLINSTGLPSRALFTPVNSMADSGD 60
QY 61 NRASDVPGLPVPMRLAASEITLNDGFEVLHDHGLDPLTLNRQIGSSVFRVETQEDGKHIA 120
DB 61 NRASDVPGLPVPMRLAASEITLNDGFEVLHDHGLDPLTLNRQIGSSVFRVETQEDGKHIA 120
QY 121 VGORNVETSVLSDOEYARLQSIDPEGKDKVFTGGRGGAGHAMVTVASDITEARQIL 180
DB 121 VGORNVETSVLSDOEYARLQSIDPEGKDKVFTGGRGGAGHAMVTVASDITEARQIL 180
QY 181 ELLEPKGTGSGKAGESKVGELRESNKAENTTETQSTSTSSLRSDPKMLALGTVA 240
DB 181 ELLEPKGTGSGKAGESKVGELRESNKAENTTETQSTSTSSLRSDPKMLALGTVA 240
QY 241 GLIGLAATGIVQALALTPEDSPPTTDPDAASATATRDQITKEAFQNPQKNVIDE 300
DB 241 GLIGLAATGIVQALALTPEDSPPTTDPDAASATATRDQITKEAFQNPQKNVIDE 300
QY 301 LGNAIPSGVLKODVANIIEQAKAAGEAKQQAENNAOAKKYDQQAQROEELKVSSG 360
DB 301 LGNAIPSGVLKODVANIIEQAKAAGEAKQQAENNAOAKKYDQQAQROEELKVSSG 360
QY 361 AGYGLSGALILGGGIGVAVTAALHRKNQVQETTTTTTTTTTTSARTVENKPNNTPAQ 420
DB 361 AGYGLSGALILGGGIGVAVTAALHRKNQVQETTTTTTTTTTTSARTVENKPNNTPAQ 420
QY 421 NVDTFGSEDTMESRRSSMASTSTFFDTSIG---GPCRIRMLMLKHKRCMIRRCRLILI 477
DB 421 NVDTFGSEDTMESRRSSMASTSTFFDTSIGTGVQNPADVKTSL-HDSQVPTNSNTSV 479
QY 478 RLFRWIGIQISVYVYTIQHPRTDNDGARLLGNPSAGIQSTYARLALSGGLRHDHMGGLT 537
DB 480 QNM---GNTDSVYVYTIQHPRTDNDGARLLGNPSAGIQSTYARLALSGGLRHDHMGGLT 536
```

```
QY 538 GGSNAVNTSNNPPAPGSHRFV 559
DB 537 GGSNAVNTSNNPPAPGSHRFV 558

RESULT 3
T25697
hypothetical protein F16F9.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C:Accession: T25697
R:Fulton, B.
submitted to the EMBL Data Library, August 1996
A:Description: The sequence of C. elegans cosmid F16F9.
A:Reference number: Z20071
A:Accession: T25697
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1229 <FUL>
A:Cross-references: EMBL:U67956; PIDN:AAB07691.1; GSPDB:GN00028; CESP:F16F9.2
A:Experimental source: strain Bristol N2; clone F16F9
C:Genetics:
A:Gene: CESP:F16F9.2
A:Map position: X
A:Introns: 35/1; 361/1; 384/1; 482/2; 517/1; 971/1; 1021/2; 1179/3

Query Match 6.0%; Score 170; DB 2; Length 1229;
Best Local Similarity 22.8%; Pred. No. 0.032;
Matches 68; Conservative 44; Mismatches 122; Indels 64; Gaps 9;

QY 167 TVASDITEARQILELLEPKGTGSGKAGESKVGELR-----ESNSGAENTTETQTST 222
DB 597 SVSTKSTTKKASTTEPTTDEPTTTESSITGKATPELSTTSEITTELKITTG 656
QY 223 SLSRSDPKMLALGTVAATGLIGLAATGIVQALALTPEDSPPTTDPDAASATATRDQ 282
DB 657 STTTEEP-----TTTAIPAEASTGII-----TTDEETTTSTTTEITSTKEIVTESA 704
QY 283 LTKEAF-----QNPDKQKNIDELGNAIPSGVLKDDV-VANIEEOAKAAGEAKQ 332
DB 705 ITQTSVSVVESSTPRQIPERKALVKNFKHLE--VLKEKKRLKKEKSTSTGSDS-- 760
QY 333 AIENNAOAKKYDQQAQROEELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQV 392
DB 761 --ETTVAENIDEVTTTEKEV-----VQTTPITTEKSTTQEE 797
QY 393 TTTTITTTTTSARTVENKPNNTPAQNVDTFGSEDTMESRRSSMASTSTFFDTS 450
DB 798 TTTTITTEKTSKTTTEKPTTSESA-----TTETTTSEPSTTSTTVDTS 844
```

```
RESULT 4
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: EMBL:U08046; PIDN:AAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2075/1;

Query Match 5.8%; Score 165.5; DB 2; Length 2232;
```

[illegible]

Query Match 5.7%; Score 162.5; DB 2; Length 1291;
Best Local Similarity 20.6%; Pred. NO. 0.095;
Matches 99; Conservative 61; Mismatches 189; Indels 131; Gaps 18

QY 5 NLGHNPVNNNSIPPAPLPSTGDG---AGGRGQLNSTGPLGSRALFTFVRNSMADSGDN 61
Db 438 NSTGNSNSNTNDSTGPGSETSTNGLVASGGAG-----GATCAAMLPTP-----SOOSTGK 488

Qy	62	RAS-----DVPGLPVFNWRLA-----ASEITLNDGDEVLDHGGPLDTLNKQ	102
Db	489	EATAAVSLLEKKLPVVVVSPLTMKELRQKGMTKYDAEMIMANAAYQQOHHH-----	539
Qy	103	IGSSVFVETQEDGHIAVGQRNGVETSVLSDQBYARLOSIDPEGDKFVFTGGRGGAG	162
Db	540	--QHFFHHHHHHHHHHHGQASTCAEATAAVQQAAMAKQPKG-----VGTTGAAC	588
Qy	163	HAMVTVASDITEARQRILLELPKGTGBSKAGESKGVGEILRESNSGAENTTETQTST	222
Db	589	NAGATTVSSVA-----AGAGSEVNGGRSTSLRKSMRVNS-----TSSSI	627
Qy	223	SSLRSDPKLWLALGTVATGLIGLAATGIVQALALTPEPDSPTTTPDDAASATETARDQ	282
Db	628	STAGADEVI-----APVVAASISLPSKAPVVLPRCKPAQMAIALHQ	670
Qy	283	LTKAEAFQNPDMQKNVIDELGNAIPSGVLVKDDVIANIEBQAKAAAGEBAKQQ-----	333
Db	671	SQORQLRRSERQEKLTD-----GESSD---TSSEQOKK----EQQODHOLQKMFS	716
Qy	334	IENNAQAQKYDEQAQRQEBELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQPVET	393
Db	717	LAESPQPEKS--EEKQEQEKRVTRNSAGRVL-----VARLATAHNN-----IA	760
Qy	394	TTTTTTTTTTSARTV---ENKPANNTPAQGNVDTPGSEDTHMESRRSSMASSTTFDRSS	450
Db	761	TTTTSSSSSNKATTITNCNHNHNSNRINHSNLSSRLSVKSRKPAPSEASSIPSSTSS	820

RESULT 6
B90835
probable tail fiber protein [imported] - Escherichia coli (strain O157:H7, subsp. C; Species: Escherichia coli)
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B90835
R:Havashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H. DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90835
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-371 <HAY>
A:Cross-references: GB:BAB35073.1; PID:gl3361114; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIND 050952
C:Genetics:
A:Gene: EC91650

Query Match	5.3%	Score 152.5;	DB 2;	Length 971;
Best Local Similarity	22.1%	Pred. No. 0.25;		
Matches	87;	Mismatched 55;	Indels 83;	Gaps 14
		Conservative		
QY	79	SEITLNDGFEVLH-----DHGFLDPLNRQIGSSVFRVETQBDGKHIAVQ-----	123	
DB	59	SVILLVEGFPSSHAGTIVYEDSOP-GTINDFLGAM-----TEDDVREALRRFELMVBE	112	
QY	124	--RNGVETSVLSQDEVARLOSIDPECKDKVFVTCGRGCGAGHANVTVASDITEARQRULE	181	
DB	113	VARN---ASAVAQNTAAKKSASDAS-----TSAREAAATHA--TDAADSARA-----	154	
QY	182	LLEPKGTGESKGAGESKGVGELRESNGAENTTQTSTSSLSRSDPKMLALGTGTVAG	241	
DB	155	-----ASTSAGQAASSAQSSASAGTASTKATEASKSAAAESSK-----SAAAT	199	
QY	242	LIGLAATGIVQALALTPEPDSPTTTDPDAAASATETATRDQLTKFAQNPDKQVNI	301	
DB	200	SAGAAKTSETNAAVSQSQSAATSASTATTKASEAASSARDASKEAKSSSETSAAS----	255	
QY	302	GNAITPSGVLDVVANTEEQAAGEEAK-QQATENNAQAOKKYDEQQAQKEELKVS	360	
DB	256	-----SASAAASSATAAGNSAKAAKTSETNAKSETTAAEQSASAAGSK--TA	301	


```

Qy 361 AGVGLSGALLGGIGVAVTAALHRKNQPVQEQTTTTTTTTTSARTVENKPNNTPAQ- 419
Db 302 AALSASAASTSGAGASATAA-----CKSAESAASASTATTKAGEATEQASAAASSASA 357
Qy 420 ---GNVDTPCEDTMSRRSSMASTSTFFDTSS 450
Db 358 AKTSETNAKASSETSAAESKTAASASSASAASSAS 391

```

RESULT 7
 C85693
 probable membrane protein of prophage CP-933X Z1918 [imported] - *Escherichia coli* [strain C]
 C:Species: *Escherichia coli*
 C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
 C:Accession: C85693
 R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
 iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, B.; Potamousis, K.; Apodaca,
 Nature 409, 529-533, 2001
 A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
 A:Reference number: A85480; MUID:21074935; PMID:11206551
 A:Accession: C85693
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-973 <STO>
 A:Cross-references: GB:AE005174; NID:gl2514847; PIDN:BAQ56007.1; GSPDB:GN00145; UWGP:Z1918
 A:Experimental source: strain O157:H7, substrain EDL933
 C:Genetics:
 A:Gene: Z1918

```

A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-770 <SCH>
A:Cross-references: EMBL:AL389901; GSPDB:GN001116; NCSP:B7F21.50
A:Experimental source: BAC clone B7F21; strain OR74A
C:Genetics:
A:Gene: NCSP:B7F21.50
A:Map position: 6
A:Introns: 117/1

Query Match          5.2%; Score 148.5; DB 2; Length 770;
Best Local Similarity 19.9%; Pred. No. 0.32;
Matches 109; Conservative 62; Mismatches 151; Indels 227; Gaps 25;

Qy      5  NLGNPN-VNNSIPPAPLPSQT---DGAGRGQLINST---GPLGSR-----ALF 48
Db      219 NSGAIPKPLPSSVPKASAIKPKTKDSGSGG---LNGAQLGGIIGAIALLIIVIAAF 274

Qy      49  TPVR---NSMADSGDNRASDVPLPVNPMRLAASEITLNDGFEVLHDHGPLDTLNRQISS 106
Db      275 LIIRRLKRVESAMESKKGTSGVHASKASTSAQAQMEQSGRFLVHRAPSOTDNASADPL 334

Qy      107 VRFVETQEDGKHI-----AVGQ---RNGVETSVVLS-----DQ 136
Db      335 MFMSETNTPGDHTTNASSLAGTQPGAHGVVGDGSGIGRHSRSGSDTTYMASPHGAGNN 394

Qy      137 EYARLOSIDP-----EGKD-----KPVFTGGRGGAGHAMVTVASDITEARQRILE 181
Db      395 NGSELASPDNRRGYFDCASPLPSSSHTTTSTTVAGHNGHGM-----RESV---442

Qy      182 LLSPKGTGSKGAGESKGVCELRESNGAENTTETOTSTSSLSRSDPKLWLALGT---237
Db      443 -----DSQSTGLGTHYSSTTRNOHHWRNQSNAELSD-----GSEITH 483

Qy      238 -VATGLIG----LAATGIV-----QALALTPEDS-----PTTTDPD 269
Db      484 GVASPLVGGSSHARGASGCTSYRYTHTHSHSHSLGVDVPELDSGFMVELPATTTTT 543

Qy      270 AAASAT-----ETATRDQITKEAFQNPDN---QKNVIDELGNAIPSGVLKDDVVA 316
Db      544 ASASALPPRNSFGLSSASRRRTTGGGSPNTSSTPQSPNTD-----586

Qy      317 NIEQAKAAGEAKQAQAIENN-----AAQAKYDSDQ 348
Db      587 STQQQQQQQQQQQQQQQQQNNLVDRSEPETAEFITRPHHQQQQQQQQQQQQQQQQQQ 646

Qy      349 -----AKRQELKVSSG-----AGYGL-----SGAL 369
Db      647 HTSLDGYGPAKQVQGTAAASGLGFVPERWDGSSPGAYPMGFPMISVPGYGLGVSSGGV 706

Qy      370 ILGGIGVA 378
Db      707 SSGGGGGGA 715

```

A:Gene: H39E23.1
A:Map position: 5

```
Query Match      5.2%; Score 148.5; DB 2; Length 1246;
Best Local Similarity 18.6%; Pred. No. 0.62; Mismatches 233; Indels 233; Gaps 25;
Matches 124; Conservative 78;

Qy 36 INSTGPGLSRALFTFVRNSMADSGDNRASDPVGLPVNPMRLAASEITLNDGFEVLHDHGP 95
Db INVSSSLGQHPAGVITREHVT-----SSSASGSSASPSRYSRSSATAT-CASITAGSAL 447

Qy 96 LDTLNQ-----IGSVFRVETQEDKHTAVGQNGVETSVVLSDOEYARLQSID 145
Db 448 ASANAQKHQSSAAPSNGSSSSRRSQNDAAATAAG-----GTVVMS----- 490

Qy 146 PEGKDKFVTGGRGAGHAMVTVASDITEARQRILELLEPKGTGSKGAGSKGVGELRE 205
Db 491 -----GTRHGGVQMAQPT-SQATISLLQPPSYKPSNTTQIAQIPLFN 535

Qy 206 SNSGAENTTQTSTSTSSLR--SDPKLWLAIGT-----VATGLIG----- 244
Db 536 RNSTA-TSSAQPSTGIGTRKIADPKGRIPLNSTAVQHRHTATGAVANNNGGIPSHRDH 594

Qy 245 -----LAATGIVQALALTPEDPSPTTDPDAASATETATRDQLTKEAFQ----- 289
Db 595 AQQQYMNQLTSSTMMSKLINKTPAAGGTAATSSSSSSSATSTA---PLQKSGSQISHAP 651

Qy 290 -----NPDNOKVNIDELGNALPS-----GVLKDDVANI-----EQAKAAGEA 329
Db 652 TEPVIREDDDDNSENQNGNVLIGGVGPQTPSPAVQVPTEDATSSSKQEQKASSTP 711

Qy 330 KQ-----QATENNAQAKKYDEQAKRQELKVSSGAGYGLSGALILGGGIGVAVTAAL 383
Db 712 KESNPIVQNLHLSLLKSLDSSNATSYETPRPGIAG----- 750

Qy 384 HRKNQPVETTTTTTTTTTSARTVENKPNANTP-----AQ 419
Db 751 -RRSEPSAATPRRRHOTVVDARHLQTPDTPRVPHFEDTTLDRQMRALYVSTASSRMT 809

Qy 420 GNDVTPGSDTWESR-----RSMASSTSTPFTDTSIGGPKCIR 458
Db 810 GVLPTPTSTNSSTSSFIPELTHVAASPDITTTTPTKSTVTTSYFRFTPS-----FR 863

Qy 459 MMLKHRCMIRRCRLILIRLFRINGIQISVVYSTIQHP-----RDTTNGA 506
Db 864 MVLIV-----LLLCDNGRLWP---SMIHQSPSNPPQMTAMESLKLSESGQTGG 911

Qy 507 RLL---GNPSAGIQSYARLALSGGLRHDMDGLTGGSNSAVNTGN----- 548
Db 912 PTVATGGPPQATSQOMRSATNTSA--NMGASSGGAATAAASATNQLSGAPSSSTGASSSQ 969

Qy 549 -NPPAPGS 555
Db 970 YHPKAPSS 977
```

RESULT 10

T18611
probable serine/threonine-specific protein kinase (SC 2.7.1.-), long splice form - Caenor
N:Contains: probable serine/threonine kinase, short splice form
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence revision 15-Oct-1999 #text_change 04-Mar-2000
C:Accession: T18611; T18610; T23144; T23143
R:McMurray, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z18997
A:Accession: T18611
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1192 <WIL1>
A:Cross-references: EMBL:Z81027; PIDN: CAB54179.1; GSPDB: GN00023; CESP: H39E23.1a
A:Experimental source: clone AH10
A:Accession: T18610

A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-487, 536-1192 <WIL2>
A:Cross-references: EMBL:Z81027; PIDN: CAB54178.1; GSPDB: GN00023; CESP: H39E23.1b
A:Experimental source: clone AH10
R:McMurray, A.
submitted to the EMBL Data Library, June 1997
A:Reference number: Z19696
A:Accession: T23144
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1192 <WIL3>
A:Cross-references: EMBL:Z96102; PIDN: CAB54263.1; GSPDB: GN00023; CESP: H39E23.1a
A:Experimental source: clone H39E23
A:Accession: T23143
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-487, 536-1192 <WIL4>
A:Cross-references: EMBL:Z96102; PIDN: CAB54262.1; GSPDB: GN00023; CESP: H39E23.1b
A:Experimental source: clone H39E23
C:Genetics:
A:Gene: CESP: H39E23.1a; CESP: H39E23.1b
A:Map position: 5
A:Introns: 1073; 333; 673; 139/2; 183/3; 451/3; 487/3; 535/3; 631/1; 825/2; 914/3; 992
C:Keywords: alternative splicing; ATP; phosphotransferase; serine/threonine-specific pro
F:1-1192/Product: probable serine/threonine-specific protein kinase, long splice form #s
F:1-487, 536-1192/Product: probable serine/threonine-specific protein kinase, short splice

Query Match 5.2%; Score 148; DB 2; Length 1192;

```
Best Local Similarity 20.6%; Pred. No. 0.62;
Matches 117; Conservative 56; Mismatches 248; Indels 148; Gaps 20;

Qy 36 INSTGPGLSRALFTFVRNSMADSGDNRASDPVGLPVNPMRLAASEITLNDGFEVLHDHGP 95
Db 505 INVSSSLGQHPAGVITREHVT-----SSSASGSSASPSRYSRSSATAT-GASITAGSAL 557

Qy 96 LDTLNQ-----IGSVFRVETQEDKHTAVGQNGVETSVVLSDOEYARLQSID 145
Db 558 ASANAQKHQSSAAPSNGSSSSRRSQNDAAATAAG-----GTVVMS----- 600

Qy 146 PEGKDKFVTGGRGAGHAMVTVASDITEARQRILELLEPKGTGSKGAGSKGVGELRE 205
Db 601 -----GTRHGGVQMAQPT-SQATISLLQPPSYKPSNTTQIAQIPLFN 645

Qy 206 SNSGAENTTQTSTSTSSLR--SDPKLWLAIGT-----VATGLIG----- 244
Db 646 RNSTA-TSSAQPSTGIGTRKIADPKGRIPLNSTAVQHRHTATGAVANNNGGIPSHRDH 704

Qy 245 -----LAATGIVQALALTPEDPSPTTDPDAASATETATRDQLTKEAFQ----- 289
Db 705 AQQQYMNQLTSSTMMSKLINKTPAAGGTAATSSSSSSSATSTA---PLQKSGSQISHAP 761

Qy 290 -----NPDNOKVNIDELGNALPS-----GVLKDDVANI-----EQAKAAGEA 329
Db 762 TEPVIREDDDDNSENQNGNVLIGGVGPQTPSPAVQVPTEDATSSSKQEQKASSTP 821

Qy 330 KQAIENNAQAKKYDEQAKRQELKVSSGAGYGLSGALILGGGIGVAVTAALHRKNQ 389
Db 822 KESKPSIHQSPSNPPQMTAMESLKLSESGQTG--GPTVATGGPPORAT--QOM 874

Qy 390 VEQTTTTTTTTTSARTVENKPNANTPAQGNVDTPGSEDWTMESRRSSMASTSTPFTDTS 449
Db 875 SRSATNSANMGASSGGAATAAASATNQLSGAPSTGASSQQYHPKAPSSSSSSST----- 929

Qy 450 SIGGPCRIRMLKHCRRMIRRCRLILIRLFRINGIQISVVYSTIQHPR---DTTNGA 506
Db 930 --NPP---HQHQLTHNA-----SPSVTPSSYQIPTSTAVNVVTGCM 965

Qy 507 RLLGNPSAGIQSYARLALSGGLRHDMDG 535
Db 966 PTSSSSSAPPNTRNQTFRHGTEDKGG 994
```

RESULT 11

G64887
probable tail fiber protein GP37 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 01-Mar-2002
C:Accession: G64887; 109189
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Cohen, A.; Rose, D.J.; Mau, B.; Shao, Y.
A:; Rose, D.J.; Mau, B.; Shao, Y.
A:Title: The complete genome sequence of Escherichia coli K-12.
Science 277, 1453-1462, 1997
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: G64887
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1122 <BLAT>
A:Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1; PID:g1787636;
A:Experimental source: strain K-12, substrain MGL655
R:Alba, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai, H.; Motomura, K.; Nakade, S.; Nakamura, Y.; Nishimoto, H.; Nishio, Y.; Oshima, T.; Saito, moto, Y.; Horiuchi, T.
DNA Res. 3, 363-377, 1996
A:Title: A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the
A:Reference number: Z16603; MUID:97251357; PMID:9097039
A:Accession: T09189
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 3-1122 <AIB>
A:Cross-references: GB:AE000234; GB:U00096; NID:g1787633; PIDN:AAC74454.1; PID:g1787636;

Query Match 5.1%; Score 146.5; DB 2; Length 1122;
Best Local Similarity 23.1%; Pred. No. 0.7;
Matches 92; Conservative 52; Mismatches 163; Indels 91; Gaps 17;
QY 79 SEITLNDGPEVLH-----DHGPLDNLNRIGSSVRFVETQEDGKHLAVGQ----- 123
Db 61 SVILLVEGPPPHAGTITVEDSQP-GTLNDFLGM-----TEDDARPEALRPFELMVEE 114
QY 124 --RNGVETSVVLSDOEYARLOSIDPEGKDKFVTGGRGAGHAMVTVASDITEARQRILE 181
Db 115 VARN--ASAVAQNTAAAKKSASDAS-----TSAREAAHTA--ADAADSARA----- 156
QY 182 LLEPKGTGSKGAGSKGKGVRENSGAENTTETQTSTSSLSRDPKMLALGTVATG 241
Db 157 -----ASTSAGQAASQAASSAGTASTKATEAKSAAAASSSK-----SAAAT 201
QY 242 LIGLAATGIVQALA-----LTPEPDSPTTDPDAASATETATRDQLTKEAFONPONQVN 297
Db 202 SAGAATSTNASSASLOSATASATATTKASEATASARDAAA-----SKEAKSSET---- 253
QY 298 IDELGNAPSGVLKDDVAVANIEQAKAGEAK-QQAIENNAQAKKYDEQQAQKEELK 356
Db 254 -----NASSS-----ASSAASSATAAGNSAKAATSETNARSSETAAGQASAAAGSK 301
QY 357 VSSGAGYGLSGALILGGGIGVAVTAALHRKNQFVQTTTTTTTTTTTTSARTVENKPANT 416
Db 302 --TAAASSAASASTSAGQASATAA-----GKSAESAASSASTATTKAGEATEQASAAAR 355
QY 417 PAQ-----GNVDTPGSEDTMESRSSMASTSTFFDTSS 450
Db 356 SASAAKTSETNAKASETSAESKTAASASSASSASSAS 393

RESULT 12

T34513
hypothetical protein ZK783.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T34513
R:Favvello, A.; Vaudin, M.
submitted to the EMBL Data Library, August 1994
A:Description: The sequence of C. elegans cosmid ZK783.
A:Reference number: Z21536
A:Accession: T34513

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
A:Residues: 1-3507 <PAV>
A:Cross-references: EMBL:U13646; PIDN:AAC24418.1; GSPDB:GN00021; CESP:ZK783.1
A:Experimental source: strain Bristol N2; clone ZK783
C:Genetics:
A:Gene: CESP:ZK783.1
A:Map position: 3
A:Introns: 14/1; 48/2; 84/1; 196/3; 303/1; 381/1; 586/1; 605/1; 1175/3; 1207/1; 1409/2;
3504/1
Query Match 5.1%; Score 144.5; DB 2; Length 3507;
Best Local Similarity 19.7%; Pred. No. 4.3;
Matches 105; Conservative 61; Mismatches 179; Indels 189; Gaps 22;
QY 27 DGAGRGQLINSTGPLGSRALFTPVR---NSMADSGDNRASDVGLPVPNPMRLAASITL 83
Db 728 EGSSGWSMTTNGITGSPSEGTIRVITLGDGPETATKGISA-PDKTGESEKTE 786
QY 84 NDGEVLHDHGPLDLNRQIGSSVRFVETQEDGKHLAVQNRNGVETSVVLSDOEYARLOS 143
Db 787 SDGE-----KLTVEKDGKE--AQSSG-----SS 808
QY 144 IDPEGKDKFVTGGRGAGHAMVTVASDITEARQRIELLEPEKGTGE-----SKGAGSKG 199
Db 809 ATSSGKKSEATSGSSSSSAKS-----GTGSEAGSSGSSSSG 846
QY 200 VGLRESNSGAENTTETQTSTSSLSRDPKMLALGTVATGLTGLAATGIVQALALTPE 259
Db 847 SGVSGSGS-----SVSTES-----GSFTSSSGSVSEATG----- 878
QY 260 PDSPTTTPDAAASATETATRDQI--TYEAFONPDNOKVNIDELGNAPSGVLKDDVVAN 317
Db 879 --STGVDGSGSKPKSPKSTEELPFTKNGEKSP----- 912
QY 318 IEEQAKAAGEE-AKQQAIEENN-----AAQAKKYDEQQAQKEE---LKVYSSGAGYGLSG 367
Db 913 SDTTGKSSSEBETTSRKPIEGSDSLTEGSGGSEWFTSGKHGFSKSVTSVSGKPTQSG 972
QY 368 ALILGGGIGVAVTAALHRKNQFVQTTTTTTTTTTTTSARTVENKPAN-----NTPAQ-- 420
Db 973 A--EGSGGPKVPGK---PGAP-EITTDGESSSTSTGDKSGGPKADKDNKNVPTGDK 1026
QY 421 --NVDTPGSEDTMESRSSMA-----STSTFFDTSSIGGPCRIEMLMLKHC 466
Db 1027 NPDTTGDSTSETSGEGQPKGSKQPPGDKGSEVKKPTSEVDGPGNLSGTKG--- 1083
QY 467 MIRCRLLILRLFRINGIQISVVYVSTIQHPRTDTTNGARLLGNPSAGIOSTY 520
Db 1084 -----SNVLPKPTDLPREGSGILTTSSGGKKNSTP 1112

RESULT 13

T33369
hypothetical protein H02F09.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 17-Mar-2000
C:Accession: T33369
R:Geisel, C.; Harmon, G.
submitted to the EMBL Data Library, July 1998
A:Description: The sequence of C. elegans cosmid H02F09.
A:Reference number: Z21330
A:Accession: T33369
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1275 <GEI>
A:Cross-references: EMBL:AF07538; PIDN:AAC64622.1; GSPDB:GN00028; CESP:H02F09.3
A:Experimental source: strain Bristol N2; clone H02F09
C:Genetics:
A:Gene: CESP:H02F09.3
A:Map position: X
A:Introns: 42/1; 78/3; 106/3; 135/2; 161/2; 224/1
C:Superfamily: Yeast glucan 1,4-alpha-glucosidase homology; glucan 1,4-alpha-glucosidase

```

165  - - - - -MVTVASDITEARQRIELLEPPKGTG----- 189
166  - - - - -VTSSDITETSHAFSSSSFFGSLSDMLLILASTEGTSGDGLRIIAVI 193
170  - - - - -TQGSIPATTQLSTTFASQKMSTVSTPTTSSIQELSTLQSQHTGSMSEISSRPQTTSVTST 815
171  - - - - -ESKSGAGE-----SKGVGELR-----ESNSG--AENTTT 214
172  - - - - -LSSSPGSGTPVQTRSVTSSSDERTNPTSSGVSNVTSPATTEVLTPTSSPSTPGNTAPRTT 875
173  - - - - -ETQTSTSTSLSDPKLMLALG-----TVAATGLIGLAATGIVQALALTPDPSPPTTT 266
174  - - - - -ETSTTTTIVLMTSLQOKLPTGSLTGSTPTTEVTITLSASSSDOVQV-----ETTQTTLL 930
175  - - - - -DPDAAA-----SATE-----TATRDQ-----LTKEAFQNPDP 292
176  - - - - -SPDATTTHAPRESSPPSTSVILTTMASTEGTSGDTGHTTAVTDQGSTPATTEISVTPS 990
177  - - - - -NOKVIMDELGNALPGVLKDDVVANIEEQAKAAGEAKQQAIEENNAQAQKDYEQQAQRQ 352
178  - - - - -TQKMS--TVSTLVTSS--TQELTSSQSGTSGMGTSSK-----PQATTP 1029
179  - - - - -EELKVSAGAGYGLSGALILGGGGIGVAVTAALHRKNQPV-----E 391
180  - - - - -TEVTTTSLSFS-----RGSIFSARNCLQTKPPLPAVVCLPDPSVPSLMHSSKPQ 1082
181  - - - - -QTTTTTTTTTTTSA-----RTVENKPNANTPAQGVNTPGSEDTHNERRRSMASTSTSTFFD 447
182  - - - - -ATTPTTEVTSTLSSFGSGTQTQTVSWETSSSGKITAPST-----SSRRTPSVATSDIFTT 1138
183  - - - - -TSSIGGCPRIIMLMKHCIRMIRCELLILIRLFRIWGIQISVVVYSTIOH---PPRDTTDN 504
184  - - - - -TDTSGNAGHTLLTGSHVITSRVASTYTLGRLL-----SIVAHQSTQRSGSTHS 1186
185  - - - - -GARLLGNPSAGIQS-----TYARLAL-----SGGLRHDWGG----- 535
186  - - - - -QSYLTESMGASSTSETSLTETATETSLCLFTWTHCDRDLLSWTSSSLTKTTKTDNRSTA 1246
187  - - - - -LTGGNSAVNTSNPPAP 553
188  - - - - -LSATSLTLPAPSTSTASRSTVPPAP 1271
189  - - - - -
190  - - - - -
191  - - - - -
192  - - - - -
193  - - - - -
194  - - - - -
195  - - - - -
196  - - - - -
197  - - - - -
198  - - - - -
199  - - - - -
200  - - - - -
201  - - - - -
202  - - - - -
203  - - - - -
204  - - - - -
205  - - - - -
206  - - - - -
207  - - - - -
208  - - - - -
209  - - - - -
210  - - - - -
211  - - - - -
212  - - - - -
213  - - - - -
214  - - - - -
215  - - - - -
216  - - - - -
217  - - - - -
218  - - - - -
219  - - - - -
220  - - - - -
221  - - - - -
222  - - - - -
223  - - - - -
224  - - - - -
225  - - - - -
226  - - - - -
227  - - - - -
228  - - - - -
229  - - - - -
230  - - - - -
231  - - - - -
232  - - - - -
233  - - - - -
234  - - - - -
235  - - - - -
236  - - - - -
237  - - - - -
238  - - - - -
239  - - - - -
240  - - - - -
241  - - - - -
242  - - - - -
243  - - - - -
244  - - - - -
245  - - - - -
246  - - - - -
247  - - - - -
248  - - - - -
249  - - - - -
250  - - - - -
251  - - - - -
252  - - - - -
253  - - - - -
254  - - - - -
255  - - - - -
256  - - - - -
257  - - - - -
258  - - - - -
259  - - - - -
260  - - - - -
261  - - - - -
262  - - - - -
263  - - - - -
264  - - - - -
265  - - - - -
266  - - - - -
267  - - - - -
268  - - - - -
269  - - - - -
270  - - - - -
271  - - - - -
272  - - - - -
273  - - - - -
274  - - - - -
275  - - - - -
276  - - - - -
277  - - - - -
278  - - - - -
279  - - - - -
280  - - - - -
281  - - - - -
282  - - - - -
283  - - - - -
284  - - - - -
285  - - - - -
286  - - - - -
287  - - - - -
288  - - - - -
289  - - - - -
290  - - - - -
291  - - - - -
292  - - - - -
293  - - - - -
294  - - - - -
295  - - - - -
296  - - - - -
297  - - - - -
298  - - - - -
299  - - - - -
300  - - - - -
301  - - - - -
302  - - - - -
303  - - - - -
304  - - - - -
305  - - - - -
306  - - - - -
307  - - - - -
308  - - - - -
309  - - - - -
310  - - - - -
311  - - - - -
312  - - - - -
313  - - - - -
314  - - - - -
315  - - - - -
316  - - - - -
317  - - - - -
318  - - - - -
319  - - - - -
320  - - - - -
321  - - - - -
322  - - - - -
323  - - - - -
324  - - - - -
325  - - - - -
326  - - - - -
327  - - - - -
328  - - - - -
329  - - - - -
330  - - - - -
331  - - - - -
332  - - - - -
333  - - - - -
334  - - - - -
335  - - - - -
336  - - - - -
337  - - - - -
338  - - - - -
339  - - - - -
340  - - - - -
341  - - - - -
342  - - - - -
343  - - - - -
344  - - - - -
345  - - - - -
346  - - - - -
347  - - - - -
348  - - - - -
349  - - - - -
350  - - - - -
351  - - - - -
352  - - - - -
353  - - - - -
354  - - - - -
355  - - - - -
356  - - - - -
357  - - - - -
358  - - - - -
359  - - - - -
360  - - - - -
361  - - - - -
362  - - - - -
363  - - - - -
364  - - - - -
365  - - - - -
366  - - - - -
367  - - - - -
368  - - - - -
369  - - - - -
370  - - - - -
371  - - - - -
372  - - - - -
373  - - - - -
374  - - - - -
375  - - - - -
376  - - - - -
377  - - - - -
378  - - - - -
379  - - - - -
380  - - - - -
381  - - - - -
382  - - - - -
383  - - - - -
384  - - - - -
385  - - - - -
386  - - - - -
387  - - - - -
388  - - - - -
389  - - - - -
390  - - - - -
391  - - - - -
392  - - - - -
393  - - - - -
394  - - - - -
395  - - - - -
396  - - - - -
397  - - - - -
398  - - - - -
399  - - - - -
400  - - - - -
401  - - - - -
402  - - - - -
403  - - - - -
404  - - - - -
405  - - - - -
406  - - - - -
407  - - - - -
408  - - - - -
409  - - - - -
410  - - - - -
411  - - - - -
412  - - - - -
413  - - - - -
414  - - - - -
415  - - - - -
416  - - - - -
417  - - - - -
418  - - - - -
419  - - - - -
420  - - - - -
421  - - - - -
422  - - - - -
423  - - - - -
424  - - - - -
425  - - - - -
426  - - - - -
427  - - - - -
428  - - - - -
429  - - - - -
430  - - - - -
431  - - - - -
432  - - - - -
433  - - - - -
434  - - - - -
435  - - - - -
436  - - - - -
437  - - - - -
438  - - - - -
439  - - - - -
440  - - - - -
441  - - - - -
442  - - - - -
443  - - - - -
444  - - - - -
445  - - - - -
446  - - - - -
447  - - - - -
448  - - - - -
449  - - - - -
450  - - - - -
451  - - - - -
452  - - - - -
453  - - - - -
454  - - - - -
455  - - - - -
456  - - - - -
457  - - - - -
458  - - - - -
459  - - - - -
460  - - - - -
461  - - - - -
462  - - - - -
463  - - - - -
464  - - - - -
465  - - - - -
466  - - - - -
467  - - - - -
468  - - - - -
469  - - - - -
470  - - - - -
471  - - - - -
472  - - - - -
473  - - - - -
474  - - - - -
475  - - - - -
476  - - - - -
477  - - - - -
478  - - - - -
479  - - - - -
480  - - - - -
481  - - - - -
482  - - - - -
483  - - - - -
484  - - - - -
485  - - - - -
486  - - - - -
487  - - - - -
488  - - - - -
489  - - - - -
490  - - - - -
491  - - - - -
492  - - - - -
493  - - - - -
494  - - - - -
495  - - - - -
496  - - - - -
497  - - - - -
498  - - - - -
499  - - - - -
500  - - - - -
501  - - - - -
502  - - - - -
503  - - - - -
504  - - - - -
505  - - - - -
506  - - - - -
507  - - - - -
508  - - - - -
509  - - - - -
510  - - - - -
511  - - - - -
512  - - - - -
513  - - - - -
514  - -
```

Db 1627 FLDNQLANGKAGNIAQALVTDHNDNLLANQSVSPALDNGAVIESQGDASSAGIVLMRPN 1686
Qy 207 NSGAENTTETQTSTSTSLRSDPKLWLA-----LGTVAATGL 242
Db 1687 NTLAGMTTIVTATLDSTGTQETLETHFVAGKAASIEMTWKDNVANNIDTNEVQVLVTDV 1746
Qy 243 IGLAATGIV-----QALALTEPDSPPTTDPDAAASATETATR-----DQLTKE 286
Db 1747 DGNAINGAVALNLTNSGMMITP---NSVTGSDGTATATLTHTLAGSLPINARIDQVSK- 1802
Qy 287 AFQNPDKVNIDELGNALPSGLKDDVVAIEEQ-----AKAAGEEAKQO 332
Db 1803 -----TINATFIADASTAQIIAGDMFIIVNDQVANGQAVNAVQARVTDSYGNPIKQO 1854
Qy 333 AIE-----NNAQAQKKYDEQAKRQBELKVSGAGYGLSGALI-----LGCGGIGVAVTAAL 383
Db 1855 TVEFVLSNNGTI-----QYELDYTSVEG-----GVMVTFNTLAGITNVATVVS 1899
Qy 384 HRKNQFVEQTNTTTTTTTTTTSAR-----TPAQGNV 422
Db 1900 SGSSRNIDTTFIADVTTAHIAASDLMWIVDDAVADNLDKNEVHARVTDKGNV 1952

Search completed: December 16, 2003, 08:56:52
Job time : 29.7392 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 26.2347 Seconds
(without alignments)
901.547 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPICNLGHNPNVNSIPPAP.....SNGAVNTSNNPPAPGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	143.5	5.0	2736	4	US-09-252-991A-30227
2	135	4.7	2137	4	US-09-134-001C-4463
3	134.5	4.7	1638	4	US-09-071-035-258
4	134.5	4.7	1638	4	US-09-071-035-262
5	134.5	4.7	1638	4	US-09-071-035-266
6	133	4.7	288	4	US-09-216-393B-341
7	133	4.7	288	4	US-09-216-393B-344
8	133	4.7	2314	4	US-09-268-347-49
9	132.5	4.6	550	4	US-09-198-452A-613
10	128	4.5	1070	4	US-09-107-532A-6735
11	127.5	4.5	558	4	US-09-071-035-268
12	126	4.4	525	4	US-09-107-532A-5095
13	125.5	4.4	461	1	US-08-186-222-2
14	125	4.4	1129	4	US-09-252-991A-29927
15	125	4.4	1702	3	US-08-296-791-5
16	125	4.4	1702	5	PCT-US95-10661A-5
17	124.5	4.4	1529	2	US-08-728-470-10
18	124.5	4.4	1529	3	US-08-719-641-10
19	124.5	4.4	1600	2	US-08-617-697-10
20	124	4.3	359	3	US-08-818-112-109
21	124	4.3	359	4	US-08-818-111-104
22	124	4.3	359	4	US-09-056-556-109
23	124	4.3	359	4	US-09-072-596-104
24	124	4.3	1619	4	US-09-328-352-7347
25	123	4.3	658	1	US-08-409-995-5
26	123	4.3	658	3	US-08-685-467-5
27	123	4.3	658	3	US-08-913-942-5

28	4.3	1098	1	US-08-409-995-2	Sequence 2, Appli
29	4.3	1098	3	US-08-685-467-2	Sequence 32, Appli
30	4.3	1098	3	US-09-377-155-32	Sequence 2, Appli
31	4.3	1098	3	US-08-913-942-2	Sequence 32, Appli
32	4.3	1098	4	US-09-669-974-32	Sequence 44, Appli
33	4.3	1098	4	US-09-268-347-44	Sequence 28, Appli
34	4.3	1220	4	US-09-206-942-28	Sequence 26, Appli
35	4.3	1226	4	US-09-206-942-26	Sequence 6, Appli
36	4.3	1848	5	PCT-US95-10661A-6	Sequence 5, Appli
37	4.3	1848	3	US-08-046-585-5	Sequence 5, Appli
38	4.3	2035	1	US-08-393-703-5	Sequence 5, Appli
39	4.3	2035	5	PCT-US93-11721-5	Sequence 5080, Ap
40	4.3	3696	4	US-09-134-001C-5080	Sequence 3, Appli
41	4.3	3992	4	US-08-621-944A-3	Sequence 48, Appli
42	4.3	1992	4	US-08-945-567D-3	Sequence 3238, A
43	4.3	2048	4	US-09-268-347-48	
44	4.3	2048	4	US-09-252-991A-32338	
45	4.2	833	4	US-09-252-991A-32338	
121	4.2	833	4	US-09-252-991A-32338	

ALIGNMENTS

RESULT 1
US-09-252-991A-30227
; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30227
; LENGTH: 2736
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
; US-09-252-991A-30227

Query Match	5.0%	Score 143.5;	DB 4;	Length 2736;
Best Local Similarity	22.1%	Pred. No. 0.0037;		
Matches 141;	Conservative 56;	Mismatches 236;	Indels 205;	Gaps 33;
QY	2	PIGNLGHNPVNSIPPAPPLPSQTDGAGRGQLINSTPLGSRALFPPVRNSMADSGDN	61	
DB	723	PIQVTDAGSGNWSFTTFLPN-----GTVNATATDAS-----GNTSAGSSVT	767	
QY	62	RASDVPLPV-NPMRLAASEITLNDGFEVLHGHPLDNLNRQIGSSVFRVETQEDGKHIA	120	
DB	768	VDSVAPATPVINE-----SNGTTLGTAEP-----GSSV-----TLTDGNGNP	805	
QY	121	VGQRN-----GVETSVLSDOEYARLQSIDPEGKDKFVFTGGRGA---GHMVTVAS	170	
DB	806	IGQVTDAGSGNWSFTTFLPN-----GTVNATATDAGN-----TSQGGSTTVVDGVAPTTPTV	860	
QY	171	DITEARQRILELLEPKGTGESKGA---GESKGVGELRESNGAENTTQTSTSSSL	225	
DB	861	NLSNGSSL-----SGTAEPGTVTLTDGNGNPVNSMADSGDN-----NWTVT-----	902	
QY	226	RSDFKMLALGTATGVLIGLAATGIVQALALTPEPSPT-----TTDPDAA	271	
DB	903	---PSPIANGTVVNVVAQDAAGNSPGASVTVDSPAAPVNVPSNGTTLGTAEPGAT	959	
QY	272	ASATETATRDQLTYEAFQNPDPNDKQVNDIELGN---AIPSGVLKDDVVANIEEQAAGEEA	329	
DB	960	VSLT-----DGNNGNPIGO-VTDAGSGNWSFTTFLPN-----GTVNATATDASPTGNTSA	1008	

QY 183 --LEPK-----GTGSKGAG-----ESKGVGELRESNGAE-NTT 214
Db 1085 GTAEPNVTBIHNEAGLVIAITGTTGAGAFITITLTGTATANEALTAIAKDAAGKSNPT 1144
QY 215 ETQTSTSSLSRDPKMLALGTVATG-----LIGLAATGIVQA 253
Db 1145 AFKTPADPADPAPVATPTVDKITGTTNGYQVVGAAEVGTTVEVRDADGTVLGMATGTDGK 1204
QY 254 LALTPEPDSPTTDD-----PDAAASATE--TATRDQUTKEAFQNPD--- 292
Db 1205 YTVLEPGKASANETITVAKNATGKESQATATTPVDLATPTIDSITGNSKGYEITGT 1264
QY 293 -NOKVNID-----ELGN--AIPSGVL-----KD----- 312
Db 1265 AEPKTTIDVRDADGTTIAATANETGQYVTLPGAVVTPGETITIIISKDGAGNESQPATA 1324
QY 313 ----DVVANIEEQAAGEAKQOAIENNAQAQKQDEQAQKQDEELKVSAGYGLSGA 368
Db 1325 VIPADVLAAPTITKVEGNKANGYTV--TGTADPNVTQFYNSEQLLA-----SGN 1374
QY 369 LILGGGIGVAVTAALHRKNQVVEQTTTTTTTTTISARTVENKPNANTPAQGNVDTPGSE 428
Db 1375 TTTGGTFSVHIAAGLATEKETL--TALTDTQGNVSPKTTFTMPADIT---GEPEIKIAA 1429
QY 429 DTWESRRSSMASTSTFFDTSSIGGPCRIRMLMKHRCMIRRCRLILILRLFIWG---I 485
Db 1430 PTV-----SSVLGTSKAGYLKGTAEPNRI--IQISNRL-----RSVIAVGATDAEGNFAI 1479
QY 486 QISVWVSTIOHPRRDITDNGARLLG-----NPSAGIQSTVARLALSGLRHDH 533
Db 1480 QLTAGQATAQOQSLATATDAGHYSTATFTMPADTPNPGGNGTGGNGTGGTGN 1539
QY 534 GGLTGGNSAVNTSNPN 550
Db 1540 GATGNGNGSNTGSPN 1556

RESULT 4

US-09-071-035-262
; Sequence 262, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 262:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-262

Query Match 4.7%; Score 134.5; DB 4; Length 1638;
Best Local Similarity 18.9%; Pred. No. 0.01;
Matches 139; Conservative 79; Mismatches 280; Indels 239; Gaps 30;

QY 6 LGHNPNVNSIPP-----APLPSTQDAGRGQLINSTGPIGS----- 44
Db 867 LAKDPAGNTSTPTTFOTPADEVVAPSVKVTGNTTQGYQVYGTGTAELGTVIERTADGTV 926
QY 45 -----RALFTPVNSMADSGNDRASDVPGLPVNPMLAASEITLND 85
Db 927 LGTATTGPTGQYVTVTLASGKATAKQTVNVVA--KNDTGLESQPTTAMTPADVTP--TIGD 983
QY 86 -----GFEVLHDHGFDLTN--RQIGSVFRVETOEDGKHIAV---GORNQVETSVVL 133
Db 984 ITGDSITGYEITGTADPNVTIEVRNPDGTTIIGTTTDDQGNFTVDLPAGAANPGDTLTVV 1043
QY 134 SDOEYARLQSIDPEGKQKVFVTCGRGGAGHAMVTVASDITEARQRIEL----- 182
Db 1044 -----GKD-----GDGNESQPTVTPEDATVAAPTVTVTGTTATGYQVT 1084
QY 183 --LEPK-----GTGSKGAG-----ESKGVGELRESNGAE-NTT 214
Db 1085 GTAEPNVTBIHNEAGLVIAITGTTGAGAFITITLTGTATANEALTAIAKDAAGKSNPT 1144
QY 215 ETQTSTSSLSRDPKMLALGTVATG-----LIGLAATGIVQA 253
Db 1145 AFKTPADPADPAPVATPTVDKITGTTNGYQVVGAAEVGTTVEVRDADGTVLGMATGTDGK 1204
QY 254 LALTPEPDSPTTDD-----PDAAASATE--TATRDQUTKEAFQNPD--- 292
Db 1205 YTVLEPGKASANETITVAKNATGKESQATATTPVDLATPTIDSITGNSKGYEITGT 1264
QY 293 -NOKVNID-----ELGN--AIPSGVL-----KD----- 312
Db 1265 AEPKTTIDVRDADGTTIAATANETGQYVTLPGAVVTPGETITIIISKDGAGNESQPATA 1324
QY 313 ----DVVANIEEQAAGEAKQOAIENNAQAQKQDEQAQKQDEELKVSAGYGLSGA 368
Db 1325 VIPADVLAAPTITKVEGNKANGYTV--TGTADPNVTQFYNSEQLLA-----SGN 1374
QY 369 LILGGGIGVAVTAALHRKNQVVEQTTTTTTTTTISARTVENKPNANTPAQGNVDTPGSE 428
Db 1375 TTTGGTFSVHIAAGLATEKETL--TALTDTQGNVSPKTTFTMPADIT---GEPEIKIAA 1429
QY 429 DTWESRRSSMASTSTFFDTSSIGGPCRIRMLMKHRCMIRRCRLILILRLFIWG---I 485
Db 1430 PTV-----SSVLGTSKAGYLKGTAEPNRI--IQISNRL-----RSVIAVGATDAEGNFAI 1479
QY 486 QISVWVSTIOHPRRDITDNGARLLG-----NPSAGIQSTVARLALSGLRHDH 533
Db 1480 QLTAGQATAQOQSLATATDAGHYSTATFTMPADTPNPGGNGTGGNGTGGTGN 1539
QY 534 GGLTGGNSAVNTSNPN 550
Db 1540 GATGNGNGSNTGSPN 1556

RESULT 5

US-09-071-035-266
; Sequence 266, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071,035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 266:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1638 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-071-035-266

Query Match 4.7%; Score 134.5; DB 4; Length 1638;
Best Local Similarity 18.9%; Pred. No. 0.01;
Matches 139; Conservative 79; Mismatches 280; Indels 239; Gaps 30;

Qy 6 LGHNPVNNISPP-----APLPSTQDAGGAGGQLINSTGPIGS----- 44
Db 867 LAKDPAGNTSTPTTFOTPADEVVAPPVSKVTGNTTQGYQVGTGAELGTTIEVRATDGTV 926
Qy 45 -----RALFPTVRNSMADSGDNRASDVPGLPVNPMLRAASEITLND 85
Db 927 LGTATTGPTGQYVTLASGKATAKQTVNVA-KNDTGLESOPTTAMPADVTP--TIGD 983
Qy 86 -----GFEVLHDHGPLDLTLN-RQIGSSVFRVETQEDGKHIAV-----GORNQVETSVVL 133
Db 984 ITGDSITGYEITGADPNTTIEVRNPDGTIIIGTTTDDQGNFTVDLPAGAANPGDITLVV 1043
Qy 134 SDQYARLQSIDPQKQKVFVTGGRGGAGHAMVTVASDITEARQRIELLE----- 182
Db 1044 -----GKD-----GDGNSQPTTEVTPEDATVAAPTIVTTVTGTTATGYQVT 1084
Qy 183 --LEPK-----GTGESKGA-----ESKGVGELRESNCAE-NTT 214
Db 1085 GTAEPNVTIEHNAGLVIAATGTTDAGAFITILPTGTATANEALTAIAKDAKESNPT 1144
Qy 215 ETQSTSTSSLRSDPKLWALGTVATG-----LIGLAATGIVQA 253
Db 1145 AFKTPADPADPVPATPTVDKITGTTNGVQVVGAAEVGTTVEVRDADGTVLGMATTTGTDGK 1204
Qy 254 LALTPEDSPPTTD-----PDAASATE--TATRDQLTKEAFQND----- 292
Db 1205 YTVTLEPGKASANETITVAKNAKTKESQPATATTPVDLATPTTIDTSITGNSKGYEITGT 1264
Qy 293 -NQKVID-----ELGN--ALPSSVL-----KD----- 312
Db 1265 AEPKTTIDVRDAGTIIAATTANETGQYVTVLPAGVTPGPTIIISKDGAGNESQPATA 1324
Qy 313 ----DVVANIEBOAKAAGEEAKQQAIAENNAQAOKKYDEQQAQKQBELKVSSGAGYLSGA 368

Db 1325 VIPADVVLAAPTITKVEGNKANGYTV--TGTADPNVTVOFYNSSQOLLA-----SGN 1374
Qy 369 LILGGGIGVAVTAALHRKNQPVQEQTTTTTTTTTTSARTVENKPAANNTPAQGNVDPGSE 428
Db 1375 TTTGGTFSVHIAAGLATEKETL--TALTDTQGNVSPKTTFTMPADIT---GEPEIKIAA 1429
Qy 429 DTMSRRSSMASTSTFTTSSIGGPCIRIMLMKHCMMIRCRLLILRLFRWG---I 485
Db 1430 PTV-----SSVLGTSKAGYLIKGTAEPNRI--IQISNRL--RSVIAVGATDAEGNFAI 1479
Qy 486 QISVVYSTTQHPPRDTTNGARLLG-----NPSAGIQSTYARLALSGGLRHDM 533
Db 1480 QLTGAQATQAQOSLLATATDAGHYSTATFTMPADPTNPGGNGNTGGNNGTGGNTGN 1539
Qy 534 GGLTGGNSAVNTSNMP 550
Db 1540 GATGGNNGSGNTGSNP 1556

RESULT 6

US-09-216-393B-341
; Sequence 341, Application US/09216393B
; Patent No. 6514694
; GENERAL INFORMATION:
; APPLICANT: Milhausen, Michael James
; TITLE OF INVENTION: TOXOPLASMA GONDII PROTEINS, NUCLEIC ACID MOLECULES, AND USES THEREOF
; FILE REFERENCE: TX-1-C2
; CURRENT APPLICATION NUMBER: US/09/216,393B
; CURRENT FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 08/994,825
; PRIOR FILING DATE: 1997-12-19
; NUMBER OF SEQ ID NOS: 366
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 341
; LENGTH: 288
; TYPE: PRT
; ORGANISM: Toxoplasma gondii
; US-09-216-393B-341

Query Match 4.7%; Score 133; DB 4; Length 288;
Best Local Similarity 18.7%; Pred. No. 0.00093;
Matches 72; Conservative 23; Mismatches 133; Indels 158; Gaps 6;

Qy 94 GPLDTLNRQIGSSVFRVETQEDGKHIAVGQ-----RNGVETSVV 132
Db 18 GPVSLARASDDSDSVETARQHMELAEADEEMHEAYDPLLEFVETFEIKKAVEEDAA 77
Qy 133 LSDQYARLQSIDPQKQKVFVTGGRGGAGHAMVTVASDITEARQRIELLE-----PK 186
Db 78 LSTDAIDRVSQFD-----LVSLLDVIREAAQAKFDLLGLRLITDIAS 118
Qy 187 GTGESKGA--GESKGVGELRESNCAENTTTQSTSTSSLRSDPKLWALGTVATGLIG 244
Db 119 GIGEGAMALMGEEAFAIRPRRSRKGKTKTTTSSSTSTT-----TTTSTTT 167
Qy 245 LAATGIVQALALTPEDSPPTTDPDAAASATEATRDQLTKEAFQNDQKNIDELGNA 304
Db 168 TTTTITT 216
Qy 305 IPSGVLKDDVIANIEOAKAAGEEAKQQAIAENNAQAOKKYDEOQAKQBELKVSSGAGY 364
Db 217 ----- 216
Qy 365 LSGALILGGGIGVAVTAALHRKNQPVQEQTTTTTTTTTTSARTVENKPAANNTPAQGNVDT 424
Db 217 -----TTTPTTTTSTTTTTTTTTTTTTTTTTTTTTTTTTTTT 248
Qy 425 PGSEDTMESRRSSMASTSTFTTSS 450
Db 249 EPTTTTITTTTTTTTTTTTTTTTTTTT 274

RESULT 7


```
TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 268:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 558 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-071-035-268

Query Match      4.5%; Score 127.5; DB 4; Length 558;
Best Local Similarity 20.3%; Pred. No. 0.0081;
Matches 116; Conservative 61; Mismatches 261; Indels 133; Gaps 20;

QY 10 PNVNNSIPPAPPL---PSQTDGAGRGQLNSTGPIQ-----SRALFTPVNSMADGDNR 62
DB 44 PNTVIEHNEAGLVIATGTTDGAGA-----FTILPTGTATANALTAIAKDAAGKSNPT 99
QY 63 ASDVPLPVNPMRLAASEITLNDGFEVLHDHPLDNLNRQIGSVFRVETQEDGKHAVG 122
DB 100 AKTPADPDAPV-----ATPTVDKITGTT-----NGYQVVG 132
QY 123 QRNGVTSVVLSDQEVARLOSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARQR---- 178
DB 133 AEGVTTVEVRDADGTVLGMATGTGDKYTVTLFPGKASANETTIVAKNATGKESQPATA 192
QY 179 --ILELLEPKGTGSKGAGSKGVGELRESNGAENTTETOTSTSTSSLSRSDPKMLMALG 236
DB 193 TTPVDLATP--TIDSITGNSKGY-----EITGTAEPKTTIDVRD----- 230
QY 237 TVATGLIGLAATGIVQALALTPBPDSPTTTTDPDAASATETATRDQLTKEAFQNPDKV 296
DB 231 -----ADGTIIAATTANETGQYTVTLPGVVTPEGIT--II 265
QY 297 NIDELGNAL--PSGLVKDDVWVNIIEQAKAAGEAKKQAIENNAQAKKYDQAKRQEE 354
DB 266 SKDGAGNESQPATAVADVPVLAAPTITTKVEGNKANGYTV--TGTADPNVTQVFNYSSEQ 323
QY 355 LKVSAGAGYGLSGLIIGGGIGVAVTAALHRKNQVPEQTITTTTTTTSARTVENKPN 414
DB 324 LLA-----SGNTTGGTFFVHIAAGLATEKELT--TALTDTQGNVSPKTTFTMPAD 373
QY 415 NTPAQGNVDTPGSDTVMESRRSSMASTSTFFDTSSICGPGCRIMLMKHKMIRRCRL 474
DB 374 IT---GEPEIKIAPTIV---SSVLGTSKAGYLKGTAEPNRI--IQISNRL-----RSV 420
QY 475 ILIRLFIRWG---IQISVVTSTQHPRPDRTDNGARLLG-----NPSAGIQST 519
DB 421 IAVGATDAEGNFAIQLTAGOATAQOSSLATATDAGHYSTATTFMTPTADPTNPGGNGNT 480
QY 520 YARLALSGGLRHDMDGGLTGGNSAVNTSNRP 550
DB 481 GGNNGTGGTNGGATGGNGNGSNTGSPN 511

RESULT 12
US-09-107-532A-5095
; Sequence 5095, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
```

```
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 5095:
SEQUENCE CHARACTERISTICS:
LENGTH: 525 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
ORGANISM: Enterococcus faecium
FEATURE:
NAME/KEY: misc feature
LOCATION: (B) LOCATION 1...525
SEQUENCE DESCRIPTION: SEQ ID NO: 5095:
US-09-107-532A-5095

Query Match      4.4%; Score 126; DB 4; Length 525;
Best Local Similarity 21.8%; Pred. No. 0.01;
Matches 77; Conservative 66; Mismatches 152; Indels 58; Gaps 13;

QY 128 ETSVLSLDQEVARLOSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARQRIILELLEPKG 187
DB 32 DSQIQQQDKIADLNKQADAQSQ-----IDALEQVSEINTQAQDLLAKQD 78
QY 188 TGSGKAGSKGVGELRESNGAENT-----TETOTSTSTSS-----LRSDPKLWLALGT 237
DB 79 TLQESAQLVKDIADLQERIEKRETDITQKAREAQVSNSTSNYIDAVLNAD-SLADAIGR 137
QY 238 VATGLIGLAATGIVQALALTPBPDSPTTTDPDAASATETATRDQLTKEAFQNP---DNQ 294
DB 138 VQ-----AMTIVKKA-----NNDLMEQKQDKKAVEKKAENDAKLKEAENQALESQ 186
QY 295 KVNIDELGNALPSGLVK-----DQVAVNIEQ-AKAAGEAK---QQAIENTNAQA 340
DB 187 KG--DLLSKQADLNVLKTSIAEQATAEDKKADLNKQKAEAEQARIREQORLAEQARQ 244
QY 341 QKKYDQQAQRQBELKVSSGAGYGLSGLIIGGGIGVAVTAALHRKNQVPEQTITTTTTT 400
DB 245 QAAQEKAEKEAREQAEAEQAQATQASSTAQSSATESSATQSSMTSESSATQSSATEEST 304
QY 401 TTTSARTVENKPNNTPAQGNVDTPGSEDVTMESR---RSSMASTSTSTFFDTSS 450
DB 305 TPES-TEESTAPESATEESTTAPESATEESTTTPESATEESTTTPESAT 356

RESULT 13
US-08-186-222-2
; Sequence 2, Application US/08186222
; Patent No. 5559007
; GENERAL INFORMATION:
; APPLICANT: Suri, Bruno
; APPLICANT: Schmitz, Albert
; TITLE OF INVENTION: Bacterial Vectors
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CIBA-GEIGY Corporation
```

STREET: 7 skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA

1052
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/186.222

Query Match 4.4%; Score 125.5; DB 1; Length 461;
Best Local Similarity 19.0%; Pred. NO. 0.0091;
Matches 84; Conservative 70; Mismatches 160; Indels 127; Gaps 16;

RESULT 14
US-09-252-991A-29927

; Sequence 29927, Application US/09252991A
 ; Patent No. 6551795
 ; GENERAL INFORMATION:
 ; APPLICANT: Marc J. Rubenfield et al.

Query Match 4.4%; Score 125; DB 4; Length 1129;
Best Local Similarity 21.4%; Pred. No. 0.041;
Matches 92; Conservative 60; Mismatches 168; Indels 110; Gaps 18;

RESULT 15
US-08-296-791-5
; Sequence 5, Application US/08296791
; Patent No. 6245337
; GENERAL INFORMATION:
; APPLICANT: St. Geme III, Joseph W.
; APPLICANT: Falkow, Stanley
; TITLE OF INVENTION: Haemophilus Adherence and Penetration
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr, Hohnbach, Test, Albritton & Herbert
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States

```

; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/296,791
; APPLICATION NUMBER: US/08/296,791
; FILING DATE: 25-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Trecartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-59941/RFT/RMS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1702 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
;
US-08-296-791-5

```

```

Query Match      4.4%; Score 125; DB 3; Length 1702;
Best Local Similarity 17.8%; Pred. No. 0.078;
Matches 87; Conservative 62; Mismatches 180; Indels 160; Gaps 17;

Qy 10 PNVNSI-----PPAPLPSTQDGAGRGQLINSTGLGSRALFTPVNRSMADSGDN 61
Db 1017 PSNNEIARVETVPVPPAPATPSETTE-----TVAENSKQES--- 1053

Qy 62 RASDPVGLPVNPMRLAASEITLNDGFEVLHDHGPLDLNRIQIGSSVFRVETQEDGKHIAV 121
Db 1054 -----KTVENKQDAITETTAQNG-EVAEEAKESVKANTQTN-----EVAQSGSETEE 1099

Qy 122 GQRNGVETSVVLSQDEYARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARQRILE 181
Db 1100 TQTEIKETAKVEKEEKAKVEK-----DEIQEAPQWASE 1141

Qy 182 LLEPKGTGSGKAGESKGVGELR-ESNSGAENTT-----ETQSTSTSSL 225
Db 1142 TSPKQAKPAPKEVSTDTKVEETQVQAPQOTQSTTVAEEATSPNSKPAEETQPSEKTN-- 1199

Qy 226 RSDPKLWALGTVATGLIGLAATGIVQALALTPEPDSPTTDDPAAASATETATRDOLTK 285
Db 1200 -AEP-----VTPVSKNOTENTDQPTEREKTKVETEKTPQPPQVAS 1241

Qy 286 EAFQNPQNKV-----NIDELGNATPSGVLKDDVVVNIIEEQAKAGEEAKQQAIE 335
Db 1242 QASPKQSQSETVQPAVLESENVTVNA-----EEVQALQQTQTSAT-VSTKQPAPE 1293

Qy 336 NNAQAQKQYDEQQAQROEELKVSSGAGYLSGALILGGIGIGVAVTAALHRKNQPVQTTT 395
Db 1294 NS-----INTGSATAITETAERKSDKPKQTETAA 1320

Qy 396 TTTTITTTTSARTV-ENKPANNTPA-----QGNVDTPGSEDTMESRRSSMASTSTFFD 447
Db 1321 STEDASQHKANTVADNSVANNSESEPKRRRRRSISQP--QETSAAETTTAASTDTEITAD 1378

Qy 448 TSSIGGPCR 456
Db 1379 NSKRSKPNR 1387

```

Search completed: December 16, 2003, 08:57:56
Job time : 29.2347 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	148.5	5.2	596	11	US-09-946-374-243	Sequence 243, App	
2	148.5	5.2	596	12	US-10-015-387A-243	Sequence 243, App	
3	148.5	5.2	596	12	US-10-063-735-100	Sequence 100, App	
4	148.5	5.2	596	12	US-10-006-130A-243	Sequence 243, App	
5	148.5	5.2	596	12	US-10-199-672-310	Sequence 310, App	
6	148.5	5.2	596	12	US-10-006-172A-243	Sequence 243, App	
7	148.5	5.2	596	12	US-10-187-749-310	Sequence 310, App	
8	148.5	5.2	596	12	US-10-184-457-310	Sequence 310, App	
9	148.5	5.2	596	12	US-10-184-642-310	Sequence 310, App	
10	148.5	5.2	596	12	US-10-196-747-310	Sequence 310, App	
11	148.5	5.2	596	12	US-10-015-392A-243	Sequence 243, App	
12	148.5	5.2	596	12	US-10-017-253A-243	Sequence 243, App	
13	148.5	5.2	596	12	US-10-173-689-310	Sequence 310, App	
14	148.5	5.2	596	12	US-10-173-690-310	Sequence 310, App	
15	148.5	5.2	596	12	US-10-173-691-310	Sequence 310, App	

[illegible]

; PRIOR FILING DATE: 1998-10-14
; PRIOR APPLICATION NUMBER: 60/104987
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105000
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105002
; PRIOR FILING DATE: 1998-10-20
; PRIOR APPLICATION NUMBER: 60/105104
; PRIOR FILING DATE: 1998-10-21
; PRIOR APPLICATION NUMBER: 60/105169
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105266
; PRIOR FILING DATE: 1998-10-22
; PRIOR APPLICATION NUMBER: 60/105693
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105694
; PRIOR FILING DATE: 1998-10-26
; PRIOR APPLICATION NUMBER: 60/105807

Query Match 5.2%; Score 148.5; DB 11; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

Qy 4 GNLGHNPNVNSIPPAPPLPSQTDGAGRGQLINSTGPGSRALFTPVNSMADSGDNRA 63
Db 13 GLLHLLEAATNS-----NETSTANTSGSVSSG-----ASTATNSG 49
Qy 64 SDVPGLPVNPMLAASEITLNDGFVHLHDGCLDNLNRQIGSSV-PRVETQEDGKHIAVG 122
Db 50 SSVTSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSSEFSSAGSISATN 107
Qy 123 QRNGVETS--VVLSDQBYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDIPEARQIL 180
Db 108 SESSTSSGASTATNSE---SSTPSSGASTVTVNSGSSVSSGASTATNSSESTVSSRAS 163
Qy 181 ELLEPKGTGSKGAGESKVGELRESNGAENTTETQSTSTSLRSDPKLMLALGTVA 240
Db 164 TATNSSESTLSSGASTATN-SDSSTSSGASTATNSSESTSSG-----ASTAT 211
Qy 241 GLIGLAATGIVQALALTPEDSPPTTDPDAAASATETATRDQLTKEAFQNPDKVNI 300
Db 212 N-----SESTVSSRASTATNSESTT---SSGASTATNSSESTSSG-----ASTAT 264
Qy 301 LG-----NAIPSGVLKDDVVANIEEQAKAAGEEAKQOAIENNAQAOKKYDEQOAKR 351
Db 265 SGASTATNSDSSTVSSGA---STATNSESTSSGAST-----ATN 302
Qy 352 QEELKVSSGAGYGL-SCALILGGIGVAVTAALHRKNQPVQETTTTTTTTSARTVEN 410
Db 303 SESSTSSGASTATNSDSSTSSGAGTATNSSESTVSSGISTVTVNSSESTPSSGANTATN 362
Qy 411 KPANTPAQGNVDTPGSEDWESRRSMASTSTFFDFTSSIGGPCRIEMLMLKHCMI 470
Db 363 SESSTSSGANTATNSSESTVSSGASTATNSSESS---TTSSG----- 401
Qy 471 CRLLILRLFRWIGIQISVVYSTIOHPRDTTNGARLLGNPSAGIOSTVARLA----- 524
Db 402 -----VSTATNSESTSSGASTATNSDSSTSSSEASTATNSSESS 441
Qy 525 -LSGGL-----RHDMGGLTGGSNSAVNTSNPPAPGS 555
Db 442 TVSSGISTVTVNSSESTSSGANTATNSGSSVTSAGS 477

RESULT 2
US-10-015-387A-243
; Sequence 243, Application US/10015387A
; Publication No. US20030135034A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnovers, Luc
; APPLICANT: Eaton, Dan I.

; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830PIC54
; CURRENT APPLICATION NUMBER: US/10/015,387A
; CURRENT FILING DATE: 2001-12-12
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 243
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-387A-243

Query Match 5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

Qy 4 GNLGHNPNVNSIPPAPPLPSQTDGAGRGQLINSTGPGSRALFTPVNSMADSGDNRA 63
Db 13 GLLHLLEAATNS-----NETSTANTSGSVSSG-----ASTATNSG 49
Qy 64 SDVPGLPVNPMLAASEITLNDGFVHLHDGCLDNLNRQIGSSV-PRVETQEDGKHIAVG 122
Db 50 SSVTSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSSEFSSAGSISATN 107
Qy 123 QRNGVETS--VVLSDQBYARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDIPEARQIL 180
Db 108 SESSTSSGASTATNSE---SSTPSSGASTVTVNSGSSVSSGASTATNSSESTVSSRAS 163
Qy 181 ELLEPKGTGSKGAGESKVGELRESNGAENTTETQSTSTSLRSDPKLMLALGTVA 240
Db 164 TATNSSESTLSSGASTATN-SDSSTSSGASTATNSSESTSSG-----ASTAT 211
Qy 241 GLIGLAATGIVQALALTPEDSPPTTDPDAAASATETATRDQLTKEAFQNPDKVNI 300
Db 212 N-----SESTVSSRASTATNSESTT---SSGASTATNSSESTSSG-----ASTAT 264
Qy 301 LG-----NAIPSGVLKDDVVANIEEQAKAAGEEAKQOAIENNAQAOKKYDEQOAKR 351
Db 265 SGASTATNSDSSTVSSGA---STATNSESTSSGAST-----ATN 302
Qy 352 QEELKVSSGAGYGL-SCALILGGIGVAVTAALHRKNQPVQETTTTTTTTSARTVEN 410
Db 303 SESSTSSGASTATNSDSSTSSGAGTATNSSESTVSSGISTVTVNSSESTPSSGANTATN 362
Qy 411 KPANTPAQGNVDTPGSEDWESRRSMASTSTFFDFTSSIGGPCRIEMLMLKHCMI 470
Db 363 SESSTSSGANTATNSSESTVSSGASTATNSSESS---TTSSG----- 401
Qy 471 CRLLILRLFRWIGIQISVVYSTIOHPRDTTNGARLLGNPSAGIOSTVARLA----- 524
Db 402 -----VSTATNSESTSSGASTATNSDSSTSSSEASTATNSSESS 441
Qy 525 -LSGGL-----RHDMGGLTGGSNSAVNTSNPPAPGS 555
Db 442 TVSSGISTVTVNSSESTSSGANTATNSGSSVTSAGS 477

RESULT 3
US-10-063-735-100
; Sequence 100, Application US/10063735
; Publication No. US2003013882A1
; GENERAL INFORMATION:

; APPLICANT: Eaton, Dan L.
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3230RIC1
; CURRENT APPLICATION NUMBER: US/10/063,735
; CURRENT FILING DATE: 2002-05-08
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 170
; SEQ ID NO 100
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-063-735-100

Query Match 5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

Qy 4 GNIGHNPVNNSTPPAPLPSTQDGAGRGQLINSTGLGSRALFTPVNSMADSGDNRA 63
Db 13 GLLHLEAATNS-----NETSANTGSSVSSG-----ASTATNSG 49
Qy 64 SDVPLPVPNMLAASEITLNDGFVHLHDGFLDPLNRQIGSSV-FRVETOEDGKHAVG 122
Db 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEPHITSSGISTATNSEFSTASSGISATN 107
Qy 123 QRNGVTS--VLSDOEYARLQSIDPEGKDFVTGGRGGAGHAMVTVASDITEARQIL 180
Db 108 SESTTSSGASTATNSE-----SSTPSSGASTVTVNSGSSVSSGASTATNSESSVSSRAS 163
Qy 181 ELLEPKGTGSKGAGESKGVGELRESNGAENTTETOTSTSTSLRSDPKLMLALGTAVT 240
Db 164 TATNSESTLSSGASTATN-SDSSTTSSGASTATNSESSITSSG-----ASTAT 211
Qy 301 LG-----NAIPSGVLKDDVVANIEQAKAAGEAKQQAIEENNAQAOKKYDEQOAKR 351
Db 265 SGASTATNSDSSVSSGA---STATNSESTTSSGAST-----ATN 302
Qy 352 QBELKVSSCAGYGL-SGALILGGGIGVAVTAALHRKNQPVQOTTTTTTTTTSARTVEN 410
Db 303 SESTTSSGASTATNSDSSSTTSSGAGTATNSESSVSSGISTVTVNSSESTPSSGANTATN 362
Qy 411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSTFFDTSSIGGFCRIRMLMLKRCMIR 470
Db 363 SESTTSSGANTATNSESTVSSGASTATNSESS---ITSSG-----401
Qy 471 CRLLILIRLFRWIGIQISVVVSTIQHPDRDTTNGARLLGNPSAGIQSTYARLA----- 524
Db 402 -----VSTATNSESTTSSGASTATNSDSSSTTSSSEASTATNSESS 441
Qy 525 -LSGGL-----RHDMGGLTGGNSAVNTSNPPAGS 555
Db 442 TVSSGISTVTVNSESTTSSGANTATNSGSSVTSAGS 477

RESULT 4

US-10-006-130A-243
; Sequence 243, Application US/10006130A
; Publication No. US20030148375A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830P1C7
; CURRENT APPLICATION NUMBER: US/10/006,130A
; CURRENT FILING DATE: 2002-03-19
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 243
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-006-130A-243

Query Match 5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

Qy 4 GNIGHNPVNNSTPPAPLPSTQDGAGRGQLINSTGLGSRALFTPVNSMADSGDNRA 63
Db 13 GLLHLEAATNS-----NETSANTGSSVSSG-----ASTATNSG 49
Qy 64 SDVPLPVPNMLAASEITLNDGFVHLHDGFLDPLNRQIGSSV-FRVETOEDGKHAVG 122
Db 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEPHITSSGISTATNSEFSTASSGISATN 107
Qy 123 QRNGVTS--VLSDOEYARLQSIDPEGKDFVTGGRGGAGHAMVTVASDITEARQIL 180
Db 108 SESTTSSGASTATNSE-----SSTPSSGASTVTVNSGSSVSSGASTATNSESSVSSRAS 163
Qy 181 ELLEPKGTGSKGAGESKGVGELRESNGAENTTETOTSTSTSLRSDPKLMLALGTAVT 240
Db 164 TATNSESTLSSGASTATN-SDSSTTSSGASTATNSESSITSSG-----ASTAT 211
Qy 241 GLIGLAATGIVQALALTPEPDSPTTTPDAAASATETATRDQLTKEAFQNPDKQKNIDE 300
Db 212 N-----SESTVSSRASATNSESTT---SSGASTATNSESTTSSGAGTATNSESS 264
Qy 301 LG-----NAIPSGVLKDDVVANIEQAKAAGEAKQQAIEENNAQAOKKYDEQOAKR 351
Db 265 SGASTATNSDSSVSSGA---STATNSESTTSSGAST-----ATN 302
Qy 352 QBELKVSSCAGYGL-SGALILGGGIGVAVTAALHRKNQPVQOTTTTTTTTTSARTVEN 410
Db 303 SESTTSSGASTATNSDSSSTTSSGAGTATNSESSVSSGISTVTVNSSESTPSSGANTATN 362
Qy 411 KPANNTPAQGNVDTPGSEDTMESRRSSMASTSTFFDTSSIGGFCRIRMLMLKRCMIR 470
Db 363 SESTTSSGANTATNSESTVSSGASTATNSESS---TTSSG-----401
Qy 471 CRLLILIRLFRWIGIQISVVVSTIQHPDRDTTNGARLLGNPSAGIQSTYARLA----- 524
Db 402 -----VSTATNSESTTSSGASTATNSDSSSTTSSSEASTATNSESS 441
Qy 525 -LSGGL-----RHDMGGLTGGNSAVNTSNPPAGS 555
Db 442 TVSSGISTVTVNSESTTSSGANTATNSGSSVTSAGS 477

RESULT 5

US-10-199-672-310

```

; Sequence 310, Application US/10199672
; Publication No. US2003014842A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/199,672
; CURRENT FILING DATE: 2002-07-18
; PRIOR APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; TYPE: PRT
; LENGTH: 596
; ORGANISM: Homo Sapien
US-10-199-672-310

Query Match 5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

Qy 4 GNIGHNPNVANSIPPAPLPSSOTDAGGRGOLINSTGPGSRALFTVPVNSMADSGDNRA 63
Db 13 GLLHLLEAATNS-----NETSANTGSSVSSG-----ASTATNSG 49

Qy 64 SDVPEGLFVNPMLAASBITLNDGPEVLHDGPLDLNROIGSSV-FRVETQEDGKHLAVG 122
Db 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFFHTSSGISTATNSSEFTASSGISATN 107

Qy 123 QRNGVETS--VVLSDQBYARLQSIDPEGKQFVFTGGRGAGHAMVTVASDITEARQIL 180
Db 108 SESSTSSGASTATNSB---SSTPSSGASTVTVNSGSSVSSGASTATNSSESTVSSRAS 163

Qy 181 ELLEPKGTGSKGAGESKGVGELRESNGAENTTETOTSTSLRSDPKLWLALGVAT 240
Db 164 TATNSSESTLSSGASTATN-SDSSTTSSGASTATNSSESTSSG-----ASTAT 211

Qy 241 GLIGLAATGIVQALALTPEDSPPTTDPDAAASATATRDOLTKEAFQNPDKKNVIDE 300
Db 212 N-----SESTVSSRASTATNSSESTT---SSGASTATNSRRTTNGAGTATNSSESTTS 264

Qy 301 LG-----NAIFPSGLVDVVVANIEEOAKAAGEEAKQOAIENNAOAKKYDEQOAKR 351
```

```

Db 265 SGASTATNSDSTVSSGA---STATNSSESTSSGAST-----ATN 302
Qy 352 QEELKVSSGAGYGL-SGALILGGIGIVAVTAALHRKNQPVQOTTTTTTTTTSARTVEN 410
Db 303 SESSTSSGASTATNSDSTSSGAGTATNSSESTVSSGISTVNSSESTPSSGANTATN 362
Qy 411 KPANNTPAQGNVDTPGSEDTMESRRSMASSTSTFFDTSSIGGFCRIMLMKHKRCMIR 470
Db 363 SESSTSSGASTATNSSESTVSSGASTATNSSEST---TTSSG----- 401
Qy 471 CRLLILRLPRFWIGIQISVVYSTIQQHPRDITDNGARLLGNPAGIQSTYARLA----- 524
Db 402 -----VSTATNSSESTSSGASTATNSDSTSSSEASTATNSSESS 441
Qy 525 -LSGGL-----RHDGMGLTGGSNAVNTSNPPAPGS 555
Db 442 TVSSGISTVNSSESTSSGANTATNSGSSVTSAGS 477

RESULT 6
US-10-006-172A-243
; Sequence 243, Application US/10006172A
; Publication No. US20030153000A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC11
; CURRENT APPLICATION NUMBER: US/10/006,172A
; CURRENT FILING DATE: 2002-03-19
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099602
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099642
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099741
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099754
; PRIOR FILING DATE: 1998-09-10
; PRIOR APPLICATION NUMBER: 60/099763
; PRIOR FILING DATE: 1998-09-10
```

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200	201	202	203	204	205	206	207	208	209	210	211	212	213	214	215	216	217	218	219	220	221	222	223	224	225	226	227	228	229	230	231	232	233	234	235	236	237	238	239	240	241	242	243	244	245	246	247	248	249	250	251	252	253	254	255	256	257	258	259	260	261	262	263	264	265	266	267	268	269	270	271	272	273	274	275	276	277	278	279	280	281	282	283	284	285	286	287	288	289	290	291	292	293	294	295	296	297	298	299	300	301	302	303	304	305	306	307	308	309	310	311	312	313	314	315	316	317	318	319	320	321	322	323	324	325	326	327	328	329	330	331	332	333	334	335	336	337	338	339	340	341	342	343	344	345	346	347	348	349	350	351	352	353	354	355	356	357	358	359	360	361	362	363	364	365	366	367	368	369	370	371	372	373	374	375	376	377	378	379	380	381	382	383	384	385	386	387	388	389	390	391	392	393	394	395	396	397	398	399	400	401	402	403	404	405	406	407	408	409	410	411	412	413	414	415	416	417	418	419	420	421	422	423	424	425	426	427	428	429	430	431	432	433	434	435	436	437	438	439	440	441	442	443	444	445	446	447	448	449	450	451	452	453	454	455	456	457	458	459	460	461	462	463	464	465	466	467	468	469	470	471	472	473	474	475	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500	501	502	503	504	505	506	507	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525
---	---	---	---	---	---	---	---	---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----


```
; PRIOR APPLICATION NUMBER: 60/106023
; PRIOR FILING DATE: 1998-10-28

Query Match
Best Local Similarity 5.2%; Score 148.5; DB 12; Length 596;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

Qy 4 GNLGHNPNVNNISPPAPPLPSQTDGAGRGQLINSTGPGSLRALFTVPRNSMADSGDNRA 63
Db 13 GLLHLEAATNS-----NETSANTGSSVSSG-----ASTATNSG 49

Qy 64 SDVPGLPVPMRLAASEITLNDGFEVLHDHCPDLTLNRQIGSSV-FRVETOEDGKHIAVG 122
Db 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTSSGISTATNSSEFSTASSGISATN 107

Qy 123 QRNGVETS--VVLSDQYARLQSIDPEKDKFVFTGGRGAGHAMVTVASDITEARQIL 180
Db 108 SESSTSSGASTATNSE--SSTPSSGASTVTVNSGSSVTSSTSGASTATNSSESVSSRAS 163

Qy 181 ELLEPKGTGSKGAGESKVGELRESNGAENTTETOTSTSTSSLRSDPKLMLALGTAT 240
Db 164 TATNSSESTLSSGASTATN-SDSSTTSSGASTATNSSESTSSG-----ASTAT 211

Qy 241 GLIGLAATGIVQALALTPEDPSPTTDPDAAASATETATRDOLTKAEAFQNDKVNIDE 300
Db 212 N-----SESSTVSSRASTATNSSESTT---SSGASTATNSERTTNSGAGTATNSSESTTS 264

Qy 301 LG-----NAIPSGVLKDDVVANIEEQAKAAGEAKQQAENNAQAOKKYDEQOAKR 351
Db 265 SGASTATNSDSSTVSSGA---STATNSSESTTSSGAST-----ATN 302

Qy 352 QBELKVSSGAGYGL-SGALILGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVEN 410
Db 303 SESSTTSSGASTATNSDSSTTSSGAGTATNSSESTVSSGISTVTVNSSESTPSSGANTATN 362

Qy 411 KPANTPAQGNVDTPGSEDIMESRRSSMASTSTFFDTSSIGGPCRLMLMLKHCIRMIR 470
Db 363 SESSTSSGANTATNSSESTVSSGASTATNSSE---TTSSG----- 401

Qy 471 CRLLILRLFRWGIQISVVYVSTIQHPRPDITDNGARLLGNPSAGIOSTVARLA----- 524
Db 402 -----VSTATNSSESTTSSGASTATNSDSSTTSSSEASTATNSSESS 441

Qy 525 -LSGGL-----RHDGMGLTGGSNAVNTSNPPAPGS 555
Db 442 TVSSGISTVTVNSSESTTSSGANTATNSGSSVTSAGS 477

RESULT 7
US-10-187-749-310
; Sequence 310, Application US/10187749
; Publication No. US20030153036A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Auelin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C1
; CURRENT APPLICATION NUMBER: US/10/187,749
; CURRENT FILING DATE: 2002-07-01
; PRIOR APPLICATION NUMBER: US/10/052,586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18

Query Match
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

Qy 4 GNLGHNPNVNNISPPAPPLPSQTDGAGRGQLINSTGPGSLRALFTVPRNSMADSGDNRA 63
Db 13 GLLHLEAATNS-----NETSANTGSSVSSG-----ASTATNSG 49

Qy 64 SDVPGLPVPMRLAASEITLNDGFEVLHDHCPDLTLNRQIGSSV-FRVETOEDGKHIAVG 122
Db 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTSSGISTATNSSEFSTASSGISATN 107

Qy 123 QRNGVETS--VVLSDQYARLQSIDPEKDKFVFTGGRGAGHAMVTVASDITEARQIL 180
Db 108 SESSTSSGASTATNSE--SSTPSSGASTVTVNSGSSVTSSTSGASTATNSSESVSSRAS 163

Qy 181 ELLEPKGTGSKGAGESKVGELRESNGAENTTETOTSTSTSSLRSDPKLMLALGTAT 240
Db 164 TATNSSESTLSSGASTATN-SDSSTTSSGASTATNSSESTTSSG-----ASTAT 211

Qy 241 GLIGLAATGIVQALALTPEDPSPTTDPDAAASATETATRDOLTKAEAFQNDKVNIDE 300
Db 212 N-----SESSTVSSRASTATNSSESTT---SSGASTATNSERTTNSGAGTATNSSESTTS 264

Qy 301 LG-----NAIPSGVLKDDVVANIEEQAKAAGEAKQQAENNAQAOKKYDEQOAKR 351
Db 265 SGASTATNSDSSTVSSGA---STATNSSESTTSSGAST-----ATN 302

Qy 352 QBELKVSSGAGYGL-SGALILGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVEN 410
Db 303 SESSTTSSGASTATNSDSSTTSSGAGTATNSSESTVSSGISTVTVNSSESTPSSGANTATN 362

Qy 411 KPANTPAQGNVDTPGSEDIMESRRSSMASTSTFFDTSSIGGPCRLMLMLKHCIRMIR 470
Db 363 SESSTSSGANTATNSSESTVSSGASTATNSSE---TTSSG----- 401

Qy 471 CRLLILRLFRWGIQISVVYVSTIQHPRPDITDNGARLLGNPSAGIOSTVARLA----- 524
Db 402 -----VSTATNSSESTTSSGASTATNSDSSTTSSSEASTATNSSESS 441

Qy 525 -LSGGL-----RHDGMGLTGGSNAVNTSNPPAPGS 555
Db 442 TVSSGISTVTVNSSESTTSSGANTATNSGSSVTSAGS 477

RESULT 8
US-10-194-457-310
; Sequence 310, Application US/10194457
; Publication No. US20030153037A1
```

```

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C296
; CURRENT APPLICATION NUMBER: US/10/194,457
; CURRENT FILING DATE: 2002-07-11
; PRIOR APPLICATION NUMBER: 10/052586
; PRIOR FILING DATE: 2002-01-15
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059266
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/063120
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063121
; PRIOR FILING DATE: 1997-10-24
; PRIOR APPLICATION NUMBER: 60/063486
; PRIOR FILING DATE: 1997-10-21
; PRIOR APPLICATION NUMBER: 60/063540
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063541
; PRIOR FILING DATE: 1997-10-28
; PRIOR APPLICATION NUMBER: 60/063544
; PRIOR FILING DATE: 1997-10-28
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-194-457-310

Query Match          5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

Qy      4  GNLGHPNVNNSIPPAPPLPSQTDGAGRGQLINSTGPLGSRALFTPVNRNSMADSGDNRA 63
Db      13  GLLHLLEAATNS-----NETSANTSGSVISSG-----ASTATNSG 49

Qy      64  SDVPGLPVNMRLAASEITLNDGFVLHDHGPDLTLNRQIGSSV-FRVETQEDGKHAVG 122
Db      50  SSVTSSGVSTATIGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSSAGSISATN 107

Qy      123  QRNGVETS--VVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARORIL 180
Db      108  SESSTSSGASTATNSE-----SSTPSSGASTVTVNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy      181  ELLEPKGTGSKAGESKGVGELRESNGAENTTETOTSTSTSLRSDPKLWLALGTAVT 240
Db      164  TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSITSSG-----ASTAT 211

Qy      241  GLIGLAATGIVQALALTPEPDSPTTDDPAAASATETATRDQLTKEAFQNPDKNOKVIDE 300
Db      212  N-----SESSTVSSRASTATNSESTT---SSGASTATNSESRRTTSSNGAGTATNSESTTS 264

Qy      301  LG-----NAIPSGVLKDDVVANIEEQAKAAGEAKQQAIIENNAQAOKKYDEQQAQR 351
Db      265  SGASTATNSDSTVSSGA---STATNSESTTSSGAST-----ATN 302

; RESULT 9
; US-10-184-642-310
; Sequence 310, Application US/10184642
; Publication No. US20030157635A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C194
; CURRENT APPLICATION NUMBER: US/10/184,642
; CURRENT FILING DATE: 2002-06-27
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
; US-10-184-642-310

Query Match          5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

Qy      4  GNLGHPNVNNSIPPAPPLPSQTDGAGRGQLINSTGPLGSRALFTPVNRNSMADSGDNRA 63
Db      13  GLLHLLEAATNS-----NETSANTSGSVISSG-----ASTATNSG 49

Qy      64  SDVPGLPVNMRLAASEITLNDGFVLHDHGPDLTLNRQIGSSV-FRVETQEDGKHAVG 122
Db      50  SSVTSSGVSTATIGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEFSSAGSISATN 107

Qy      123  QRNGVETS--VVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEARORIL 180
Db      108  SESSTSSGASTATNSE-----SSTPSSGASTVTVNSGSSVTSSGASTATNSESSTVSSRAS 163

Qy      181  ELLEPKGTGSKAGESKGVGELRESNGAENTTETOTSTSTSLRSDPKLWLALGTAVT 240
Db      164  TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESSITSSG-----ASTAT 211

Qy      241  GLIGLAATGIVQALALTPEPDSPTTDDPAAASATETATRDQLTKEAFQNPDKNOKVIDE 300
Db      212  N-----SESSTVSSRASTATNSESTT---SSGASTATNSESRRTTSSNGAGTATNSESTTS 264

Qy      301  LG-----NAIPSGVLKDDVVANIEEQAKAAGEAKQQAIIENNAQAOKKYDEQQAQR 351
Db      265  SGASTATNSDSTVSSGA---STATNSESTTSSGAST-----ATN 302
```


Matches	102;	Conservative	84;	Mismatches	255;	Indels	135;	Gaps	197
Qy	4	GNLGHNPVNNSIPPAPLP	PSQTGDAGGRGOLINSTGLSRALFTPV	RNMSADSGDNRA	63				
Dd	13	GLLLHLEAATNS-----	NETSTANTGGSSVISSG-----	ASTATNSG	49				
Qy	64	SDVPEGLPWNWRLLAASEILTNDGFEVLHDHG	PDLTLNRQIGSSV-FRVETQEDGKHIAVG	122					
Dd	50	SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGI	STATNSEPFSTASSGISIATN	107					
Qy	123	QRNGVEYS--VVLDQEVARLQSIDPECKDFVF	TGRRGGAGHAMVTVASDITEARQKIL	180					
Dd	108	SESSTTSSGASTATNSE-----SSTPSSGASVTNS	GGSVTSSGASTATNSESTTVSSRAS	163					
Qy	181	ELLEPKGTGESKGAGESKGVGELRESNGAENTTET	OTSTLSLRSPDKLWLALGTVAT	240					
Dd	164	TATNSESTTSSLGASTATN-SDSSTTSSGASTATNS	ESSTTSSG-----ASTAT	211					
Qy	241	GLIGLATGIVOALALTPPEPDSPTTTDDPAASA	TATATROLTKEARQNPDKNVNIDE	300					
Dd	212	N-----SESSTVSSRASTATNSESTT---SSG	ASTATNSESTTNSGAGTATNSESTTS	264					
Qy	301	LG-----NAIPSGVLKDDEVANIEEQAKAE	EAKQAIIENNAOAKKYDQQAQR	351					
Dd	265	SGASTATNSDSTVSSGA---STATNSESTTSSG	AST---	302					
Qy	352	QEELKVSSGAGYGL-SGALLGGIGVAVTAALHR	KNQFVEQTTTTTTTTTTSARTVEN	410					
Dd	303	SESSTTSSGASTATNSDSTTSSGAGTATNSEST	TVSSGITVNTNSESTPSSGANTATN	362					
Qy	411	KPANNTPAQGNVDTPGSBDTWESRRSSMAWST	SSTFFDTSSICGPCRIRMLMKHRCMIRR	470					
Dd	363	SESSTTSSGANTATNSESTTVSSGASTATNSESS	--TTSSG-----	401					
Qy	471	CRLLILIRLFIWGIIQSWVYSTIOHPPRDITDNC	ARLLGNPSAGIQSTYARLA-----	524					
Dd	402	-----VSTATNSESTTSSGASTATNSDST	TSSEASTATNSESS	441					
Qy	525	-LSGGL-----RHMGLTGGSNSAVNTNNPP	APGS	555					
Dd	442	TVSSGITVNTNSESTTSSGANTATNSGSSVT	SAGS	477					

RESULT 12
US-10-017-253A-243
; Sequence 243, Application US/10017253A
; Publication No. US20030166055A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Deenoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; TITLE OF INVENTION: Acids Encoding the Same
; FILE REFERENCE: P2830PIC62
; CURRENT APPLICATION NUMBER: US/10/017,253A
; CURRENT FILING DATE: 2001-12-13
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01

```

; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 243
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-017-253A-243

Query Match      5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19

Qy 4 GNIGHNPNVANSIPPAPPLPSQTDGAGRGQLINSTGLSGRALFPPVRNSMASSGDNRA 63
Db 13 GLLHLLEAATNS-----NETSTSA NTGSSVSSG-----ASTATNSG 49

Qy 64 SDVPGLPVNPMLRAASBITLNDGFEVLHDHCPLDTLNRQIGSSV-PRVETQEDGKHVAVG 122
Db 50 SSVTSSGVSSTATIGSSVTEN-GVSIIV-TNSEFHTTSGIGSTATNSSEFSTASSGISATN 107

Qy 123 QRNGVETS--VVLSDQVYARLQSIDPBGKDFVFTGGRGGAGHAMVTVASDITEARQRIL 180
Db 108 SESSTTSSGASTATNSE-----SSTPSSGASIVTNSGSSVTSSTSGASTATNSSESTVSSRAS 163

Qy 181 ELLEPKGTGSKGAGESKGVEELRESNGSAGNTEETOTSTSTLSRSDPKLWLALGTVAT 240
Db 164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSSESTTSSG-----ASTAT 211

Qy 241 GLIGLATGIVQALALTEPDPSPPTTDPDAAASATETATRDOLTKEAFQNPDNOKNWIDE 300
Db 212 N-----SSSSTVSSRASTATNSSESTT---SSGASTATNSSERTTNGAGTATNSSESTTS 264

Qy 301 LG-----NAIPSGVLKDDVVANIEEQAKAAGEEAKQAIENNAQAQKYIDEQQAQR 351
Db 265 SGASTATNSDSSTVSSGA---STATNSSESTTSSGAST-----ATN 302

Qy 352 QEELKVSAGYGL-SGALILGGGIVAVTAALHRKNQPVQETTTTTTTTTTSARTVEN 410
Db 303 SESSTTSSGASTATNSDSSTTSSGAGTATNSSESTVSSGIGTIVTNSSESTTSSGANTATN 362

Qy 411 KPANNTPAQGVNDTPGSEDITWESRRSSMASTSSFFDTSSIGGPCIRLMLKHCMMIR 470
Db 363 SESSTTSSGANTATNSSESTVSSGASTATNSSESS---TTSSG----- 401

Qy 471 CRLLILRLPRIWGQISVYVSTTIOHPDRDITDNGARLLGNPSAGIQSTYARLA----- 524
Db 402 -----VSTATNSSESTTSSGASTATNSDSSTTSSSEASTATNSSESS 441

Qy 525 -LSGL-----RHDMGGLTGGNSAVNTSNPPAPGS 555
Db 442 TVSSGIGTIVTNSSESTTSSGANTATNSGSSVTSAGS 477

```

RESULT 13
US-10-173-689-310 ; Sequence 310, Application US/10173689
; Publication No. US200301661041
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.

APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C10
CURRENT APPLICATION NUMBER: US/10/173,689
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 310
LENGTH: 596
TYPE: PRT
ORGANISM: Homo Sapien
US-10-173-689-310

Query Match 5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;
QY 4 GNIGHNPNVNSIPPAPPLPSQTDGAGRGQLINSTGPGSRALFTPVNSMADSGNRA 63
Db 13 GLLHLEAATNS-----NETSANTSGSVISSG-----ASTATNSG 49
QY 64 SDVPGLPVNMRLAASEITLNDGFEVLHDHGLDPLTNRQIGSSV-FRVETQEDCKHIAVG 122
Db 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSESTASSGISIATN 107
QY 123 QRNGVETS--VVLSDQYARLQSIDPEKDKFVFTGGRGGAGHAMVTVASDITEARQIL 180
Db 108 SESSTSSGASTATNSE-----SSTPSSGASTVTVNSGSSVTSSGASTATNSESTVSSRAS 163
QY 181 ELLEPKGTGSKAGESKGVGELRESNGAENTTETOTSTSTSLRSDPKLWLAIGTVAT 240
Db 164 TATNSESTSSGASTATN-SDSSTTSSGASTATNSESTSSG-----ASTAT 211
QY 241 GLIGLAATGIVQALALTPEPDSPTTDPDAAASATETATRDQLTKEAFQNPDKNOKVIDE 300
Db 212 N-----SESSTVSSRASTATNSESTT---SSGASTATNSESTTSSGAGTATNSESTTS 264
QY 301 LG-----NAIPSGVLKDDVVANIEEQAKAGEEAKQQAIEENNAQAQKYDEQOAKR 351
Db 265 SGASTATNSDSTVSSGA---STATNSESTTSSGAST-----ATN 302
QY 352 QBELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVQTTTTTTTTTSARTVEN 410
Db 303 SESSTSSGASTATNSDSTTSSGAGTATNSESTVSSGISTVTVNSSTPSSGANTATN 362
QY 411 KPANNTPAQGNVDTPGSEDTMESRRSMASTSTFFDPTSSIGGPCRIIRMLKHCIRMIR 470
Db 363 SESSTSSGANTATNSESTVSSGASTATNSESS---TTSSG-----401
QY 471 CRLLILRLFRWGIQISVVYVSTIQHPDPTDNGARLLGNPSAGIOSTYARLA-----524
Db 402 -----VSTATNSESTTSSGASTATNSDSTTSSSEASTATNSESS 441
QY 525 -LSGGL-----RHDGMGLTGGNSAVNTSNPPAPGS 555
Db 442 TVSSGISTVTVNSSTSSGANTATNSGSSVTSSAGS 477

RESULT 14

US-10-173-690-310
; Sequence 310, Application US/10173690
; Publication No. US20030166105A1
; GENERAL INFORMATION:

APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
FILE REFERENCE: P3430R1C9
CURRENT APPLICATION NUMBER: US/10/173,690
CURRENT FILING DATE: 2002-06-17
Prior Application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 310
LENGTH: 596
TYPE: PRT
ORGANISM: Homo Sapien
US-10-173-690-310

Query Match 5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;
QY 4 GNIGHNPNVNSIPPAPPLPSQTDGAGRGQLINSTGPGSRALFTPVNSMADSGNRA 63
Db 13 GLLHLEAATNS-----NETSANTSGSVISSG-----ASTATNSG 49
QY 64 SDVPGLPVNMRLAASEITLNDGFEVLHDHGLDPLTNRQIGSSV-FRVETQEDCKHIAVG 122
Db 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSESTASSGISIATN 107
QY 123 QRNGVETS--VVLSDQYARLQSIDPEKDKFVFTGGRGGAGHAMVTVASDITEARQIL 180
Db 108 SESSTSSGASTATNSE-----SSTPSSGASTVTVNSGSSVTSSGASTATNSESTVSSRAS 163
QY 181 ELLEPKGTGSKAGESKGVGELRESNGAENTTETOTSTSTSLRSDPKLWLAIGTVAT 240
Db 164 TATNSESTSSGASTATN-SDSSTTSSGASTATNSESTSSG-----ASTAT 211
QY 241 GLIGLAATGIVQALALTPEPDSPTTDPDAAASATETATRDQLTKEAFQNPDKNOKVIDE 300
Db 212 N-----SESSTVSSRASTATNSESTT---SSGASTATNSESTTSSGAGTATNSESTTS 264
QY 301 LG-----NAIPSGVLKDDVVANIEEQAKAGEEAKQQAIEENNAQAQKYDEQOAKR 351
Db 265 SGASTATNSDSTVSSGA---STATNSESTTSSGAST-----ATN 302
QY 352 QBELKVSSGAGYGL-SGALILGGGIGVAVTAALHRKNQPVQTTTTTTTTTSARTVEN 410
Db 303 SESSTSSGASTATNSDSTTSSGAGTATNSESTVSSGISTVTVNSSTPSSGANTATN 362
QY 411 KPANNTPAQGNVDTPGSEDTMESRRSMASTSTFFDPTSSIGGPCRIIRMLKHCIRMIR 470
Db 363 SESSTSSGANTATNSESTVSSGASTATNSESS---TTSSG-----401
QY 471 CRLLILRLFRWGIQISVVYVSTIQHPDPTDNGARLLGNPSAGIOSTYARLA-----524
Db 402 -----VSTATNSESTTSSGASTATNSDSTTSSSEASTATNSESS 441
QY 525 -LSGGL-----RHDGMGLTGGNSAVNTSNPPAPGS 555
Db 442 TVSSGISTVTVNSSTSSGANTATNSGSSVTSSAGS 477

RESULT 15

US-10-173-691-310
; Sequence 310, Application US/10173691
; Publication No. US20030166106A1

; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C13
; CURRENT APPLICATION NUMBER: US/10/173,691
; CURRENT FILING DATE: 2002-06-17
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 310
; LENGTH: 596
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-173-691-310

Query Match 5.2%; Score 148.5; DB 12; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0046;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

QY 4 GNLGNPNVNNISIPAPLPDQDGGGQGLINSTGLGRALFTPVNSMADSGDNRA 63
Db 13 GILLHLEATNS-----NETSTANTGSSVSSG-----ASTATNSG 49
QY 64 SDVPLGPNMPLAASEITLNDGFVLHDHGLDPLNRLQIGSSV-FRVETQEDGKHIAVG 122
Db 50 SSVTSSGVSTATIGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSEPSTASSGISATN 107
QY 123 QNNGVETS--VVLSDQEVARLQSDPEGKDFVFTGGRGAGHAMVTVASDITEARQRL 180
Db 108 SESSTSSGASTATNSE----SSTPSSGASTVNSGSSVTSSTSGASTATNSESTVSSRAS 163
QY 181 ELLEPKGTGSGKAGESKGVGELRESNGAENTTETQTSTSSLRSDPKLWLALGVAT 240
Db 164 TATNSESTLSSGASTATN-SDSSTSSGASTATNSESTTSSG-----ASTAT 211
QY 241 GLIGLAATCIQAALATPEPDSPTTDPDAASATETATRDQLTKEAFQNPNDOKVNIDE 300
Db 212 N-----SESTVSSRSTATNSESTT---SSGASTATNSESTTNGAGTATNSESTTS 264
QY 301 LG-----NAIPSGVLKDDVVANIIEQAKAAGEAKQAIAENNAQAQKYDEQAKR 351
Db 265 SGASTATNDSSTVSSGA---STATNSESTTSSGAST-----ATN 302
QY 352 QBELKVSAGAGYGL-SGALILGGIGVAVTAALHRKNQPVQOTTTTTTTTTSARTVEN 410
Db 303 SESSTSSGASTATNSDSSTSSGAGTATNSESTVSSGISTVNSSESSTPSSGANTATN 362
QY 411 KPANTPAGNVDTGSEDTMESRSMASSTSTFTDSSIGGPCRIRMLMLKHCMIIR 470
Db 363 SESSTSSGANTATNSESTVSSGASTATNSESS---TTSSG----- 401
QY 471 CRLLILRLFRWGIQISVVYGTIOHPDRDITDNGARLIGNPSAGIQSTYARLA----- 524
Db 402 -----VSTATNSESTTSSGASTATNSDSSTTSEASTATNSESS 441
QY 525 -LSGGL-----RHDMGGLTCGSNAVNNTNNPPAGS 555
Db 442 TVSSGISTVTNGESSTTSSGANTATNSGSSVTSAGS 477

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 53.9828 Seconds
(without alignments)
1643.636 Million cell updates/sec

Title: US-09-189-415A-4
Perfect score: 2851
Sequence: 1 MPTGNLGHNPVNSIPPAP.....SNAVNTSNNPPAGSHRFV 559

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES.

Result No.	Score	Query Match	Length	ID	Description
1	2851	100.0	559	20	AAV06221
2	1498	52.5	549	20	AAV06220
3	406	14.2	107	21	AA20576
4	162.5	5.7	1300	22	AB268075
5	161	5.6	2586	22	AB268078
6	152.5	5.3	1412	22	AB268093
7	149.5	5.2	676	22	AB262247
8	148.5	5.2	596	21	AAV99408
9	148.5	5.2	596	22	AAU29178

10	148.5	5.2	596	22	AA887575	Human PRO1342. HO
11	148.5	5.2	596	22	AA866157	Protein of the inv
12	148.5	5.2	596	23	ABG95900	Human secreted/tra
13	148.5	5.2	596	24	ABU71266	Human PRO1342 prot
14	148.5	5.2	596	24	ABU71555	Human secreted pol
15	148.5	5.2	596	24	ABU72001	Novel human secret
16	148.5	5.2	596	24	ABU72158	Human PRO polypept
17	148.5	5.2	596	24	ABU65723	Human secreted/tra
18	148.5	5.2	596	24	ABU66056	Novel human secret
19	148.5	5.2	596	24	ABU67560	Human secreted/tra
20	148.5	5.2	596	24	ABU65418	Human PRO polypept
21	148.5	5.2	596	24	ABU58554	Human PRO polypept
22	148.5	5.2	596	24	ABU56090	Human secreted/tra
23	148.5	5.2	596	24	ABU57085	Human PRO polypept
24	148.5	5.2	596	24	ABU10664	Human secreted/tra
25	145	5.1	332	22	AB267152	Drosophila melanog
26	145	5.1	1870	24	ABJ19019	Pathogen specific
27	142.5	5.0	914	22	AB266394	Drosophila melanog
28	142.5	5.0	1026	22	AB261775	Drosophila melanog
29	141.5	5.0	1086	22	AB269356	Drosophila melanog
30	141.5	5.0	1959	12	AA10562	Mutant protease (d
31	141.5	5.0	1962	12	AA10560	Mutant protease (K
32	141.5	5.0	1962	12	AA10561	Mutant protease (N
33	141.5	5.0	1962	12	AA10557	Mutant protease (A
34	141.5	5.0	1962	12	AA10558	Mutant protease (A
35	141.5	5.0	1962	12	AA10559	Mutant protease (A
36	141.5	5.0	1962	12	AA10563	Mutant protease (K
37	141.5	5.0	1968	12	AA10941	Mutant protease (d
38	141.5	5.0	1974	12	AA10940	Mutant protease (d
39	140.5	4.9	875	22	AB271072	Drosophila melanog
40	140.5	4.9	1026	21	AA183025	Staufen protein of
41	138	4.8	688	22	AB250180	Human liver peptid
42	138	4.8	688	22	AB250137	Peptide #2788 enco
43	138	4.8	688	22	AB250137	Peptide #2813 enco
44	138	4.8	688	22	AB20749	Protein #2748 enco
45	138	4.8	688	22	AA256138	Human brain expres

ALIGNMENTS

RESULT 1
AAV06221
ID AAV06221 standard; Protein; 559 AA.

XX AC AAV06221;

XX DT 16-AUG-1999 (first entry)

XX DE EHEC E. coli translocated intimin receptor (Tir).

XX DE Tir; translocated intimin receptor; Hp90; enterohaemorrhagic;
XX KW EHEC; infection; diagnosis; vaccine.

XX OS Escherichia coli.

XX FH Key Location/Qualifiers

XX FT Misc-difference 453 /note= "encoded by codon of 1 apparent nucleotide,
XX FT causing frameshift in the DNA sequence"

XX PN WO924576-A1.

XX PD 20-MAY-1999.

XX PF 10-NOV-1998; 98WO-CA01042.

XX PR 12-NOV-1997; 97US-0065130.

XX PA (UYBR-) UNIV BRITISH COLUMBIA.

XX PI Devinney R, Finlay BB, Kenny B, Stein M;

XX XX

DR WPI; 1999-337712/28.
XX N-PSDB; AAX58859.
XX New translocated intimin receptor useful for treating infection by
PT enteropathogenic or enterohemorrhagic Escherichia coli
XX
PS Claim 7; Page 55-58; 91pp; English.
XX
CC The present sequence represents Tir, a novel translocated intimin
CC receptor (formerly termed Hp90) from an enterohaemorrhagic
CC Escherichia coli (EHEC) strain. The sequence was deduced from an
CC isolated tir polynucleotide (see AAX58859). Tir proteins are
CC secreted by attaching and effacing pathogens such as EHEC and EPEC
CC (see AAY06220) E. coli. The bacterial pathogens insert their own
CC receptors into mammalian cell surfaces, to which the pathogen then
CC adheres to trigger additional host signaling events and actin
CC nucleation. Diagnosis of disease caused by pathogenic E. coli can
CC be performed by use of antibodies that bind to Tir to detect the
CC protein or the use of nucleic acid probes for detection of nucleic
CC acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
CC peptides, a recombinant method for producing recombinant Tir,
CC antibodies which bind to Tir, and a kit for the detection of
CC Tir-producing E. coli are provided. A method of immunising a host
CC with Tir to induce a protective immune response is also provided.
CC In addition, Tir fusion proteins can be used in attenuated E. coli
CC to induce a cell-mediated immune response to other polypeptides,
CC e.g. antigens. A method for screening for compounds which
CC interfere with the binding of bacterial pathogens to their
CC receptors is further provided.
XX
SQ Sequence 559 AA;
Query Match 100.0%; Score 2851; DB 20; Length 559;
Best Local Similarity 100.0%; Pred. No. 4.4e-212;
Matches 559; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MPIGNLGNPNVNNISIPAPPLPSQTDGAGRGQLINSTGLGRALFTPVNSMADSGD 60
Db 1 MPIGNLGNPNVNNISIPAPPLPSQTDGAGRGQLINSTGLGRALFTPVNSMADSGD 60
Qy 61 NRASDVPGLPVNRLAASEITLNDGFEVLHDHGLPLTLNRQIGSSVFRVETQDGGHIA 120
Db 61 NRASDVPGLPVNRLAASEITLNDGFEVLHDHGLPLTLNRQIGSSVFRVETQDGGHIA 120
Qy 121 VQQRNGVETSVLSDQEVRLQSIDPEGKDFVFTGGGAGHAWTVASDITEARQIL 180
Db 121 VQQRNGVETSVLSDQEVRLQSIDPEGKDFVFTGGGAGHAWTVASDITEARQIL 180
Qy 181 ELLEPKGTGSGKAGESKGVGELRESNSGAENTTETQSTSTSLRSDPKLWALGTAT 240
Db 181 ELLEPKGTGSGKAGESKGVGELRESNSGAENTTETQSTSTSLRSDPKLWALGTAT 240
Qy 241 GLIGLATGIQVQALATPEPSPPTTDDAASATETATROLTKEAFQPNQKNVIDE 300
Db 241 GLIGLATGIQVQALATPEPSPPTTDDAASATETATROLTKEAFQPNQKNVIDE 300
Qy 301 LGNAIPSVLKDQVVAETEEAKAGBEAKQQAENNAQAOKKYDEQAKROELKVSSG 360
Db 301 LGNAIPSVLKDQVVAETEEAKAGBEAKQQAENNAQAOKKYDEQAKROELKVSSG 360
Qy 361 AGYGLSGALILGGGIGVAVTAALHRKNQFVQTTTTTTTTTTSARTVENKPNANTPAQG 420
Db 361 AGYGLSGALILGGGIGVAVTAALHRKNQFVQTTTTTTTTTTSARTVENKPNANTPAQG 420
Qy 421 NVDTFGSDTWESESSMASTSSPTTSSITGGPCIRMLKRCIMRRCRLILRLF 480
Db 421 NVDTFGSDTWESESSMASTSSPTTSSITGGPCIRMLKRCIMRRCRLILRLF 480
Qy 481 RIWGIQISVYVSTIQHPDDTTCARLLGNPSAGIQSTYARLALSGLRHDMGLTGGG 540
Db 481 RIWGIQISVYVSTIQHPDDTTCARLLGNPSAGIQSTYARLALSGLRHDMGLTGGG 540
Qy 541 NSAVNTSNNPPAPGSHRFV 559

Db 541 NSAVNTSNNPPAPGSHRFV 559
RESULT 2
AAY06220
ID AAY06220 standard; Protein; 549 AA.
XX AAY06220;
XX AC
XX 16-AUG-1999 (first entry)
XX DE EPEC E. coli translocated intimin receptor (Tir).
XX Tir; translocated intimin receptor; Hp90; enteropathogenic;
XX EPEC; infection; diagnosis; vaccine.
XX OS Escherichia coli.
XX Key Location/Qualifiers
FH Domain 234..253 /note= "putative transmembrane domain"
FT Domain 364..386 /note= "putative transmembrane domain"
FT Misc-difference 180 /note= "encoded by AAA"
FT Misc-difference 314 /note= "given as Xaa in the specification; Lys
is deduced from the DNA sequence"
XX WO9924576-A1.
XX 20-MAY-1999.
XX 10-NOV-1998; 98WO-CA01042.
XX 12-NOV-1997; 97US-0065130.
XX (UVBR-) UNIV BRITISH COLUMBIA.
XX Devinney R, Finlay BB, Kenny B, Stein M;
XX WPI; 1999-337712/28.
XX N-PSDB; AAX58859.
XX New translocated intimin receptor useful for treating infection by
XX enteropathogenic or enterohemorrhagic Escherichia coli
XX Claim 6; Page 55-58; 91pp; English.
XX The present sequence represents Tir, a novel translocated intimin
XX receptor (formerly termed Hp90) from an enteropathogenic
XX Escherichia coli (EPEC) strain. The sequence was deduced from an
XX isolated tir polynucleotide (see AAX58858). Tir proteins are
XX secreted by attaching and effacing pathogens such as EPEC and EHEC
XX (see AAY06221) E. coli. The bacterial pathogens insert their own
XX receptors into mammalian cell surfaces, to which the pathogen then
XX adheres to trigger additional host signaling events and actin
XX nucleation. Diagnosis of disease caused by pathogenic E. coli can
XX be performed by use of antibodies that bind to Tir to detect the
XX protein or the use of nucleic acid probes for detection of nucleic
XX acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
XX peptides, a recombinant method for producing recombinant Tir,
XX antibodies which bind to Tir, and a kit for the detection of
XX Tir-producing E. coli are provided. A method of immunising a host
XX with Tir to induce a protective immune response is also provided.
XX In addition, Tir fusion proteins can be used in attenuated E. coli
XX to induce a cell-mediated immune response to other polypeptides,
XX e.g. antigens. A method for screening for compounds which
XX interfere with the binding of bacterial pathogens to their
XX receptors is further provided.
XX Sequence 549 AA;
SQ

XX	WPI; 2000-499357/44.
XX	Screening for inhibitors of intimin binding to eukaryotic cells, for
PT	use in diagnosing, preventing and treating bacterial infections,
PT	especially Escherichia coli O157:H7
XX	Claim 8; Page 76; 96pp; English.
XX	The present invention describes a method of screening for an inhibitor
CC	of intimin binding to eukaryotic cells. The method comprises exposing an
CC	intimin polypeptide having a Tir-independent cell binding activity to
CC	test agents, and obtaining an inhibitor based on its ability to bind the
CC	polypeptide. The inhibitors are used in the prevention, treatment and/or
CC	diagnosis of bacterial infections, preferably by enteropathic and/or
CC	enterohaemorrhagic Escherichia coli, Shiga toxinigenic E. coli, Hafnia
CC	alvei or Citrobacter freundii, or especially E. coli O157:H7. The
CC	infections cause a histopathological effect known as attachment and
CC	effacement on intestinal epithelial cells. The inhibitors can be used
CC	to produce food supplements or additives, especially where the food is
CC	a milk substitute. The method can be used to sort cells based on their
CC	ability to bind to a Tir independent cell binding domain of an intimin
CC	polypeptide. Polypeptides having Tir-independent intimin binding
CC	activity can be used to produce a vaccine against a bacterial disease.
CC	The present sequence represents a specifically claimed intimin C-terminal
CC	Tir binding domain amino acid sequence, for use in the method of the
CC	present invention.
XX	Sequence 107 AA;
QY	Query Match 14.2%; Score 406; DB 21; Length 107;
DB	Best Local Similarity 72.0%; Pred. No. 8.7e-24;
XX	Matches 77; Conservative 18; Mismatches 12; Indels 0; Gaps 0;
QY	252 QALALTEPDSPPTTDPDAAASATATATRDOLTKEAFQNDPNQKVNIDELGNAIPSGVLK 311
DB	1 QALALTEPDPTTTDPDQANAASATKDQLTQEAFFKNPENQKVNIDANGNAIPSGELK 60
QY	312 DVVAVNIEQAKAAGEAKQQAENNAQAKKYDEQQAQRYEDQHARRQBELKVS 358
DB	61 DDIVEQIAQAKEAGEVARQQAQVESAQAQRYEDQHARRQBELQLS 107
RESULT 4	
ABB68075	
ID	ABB68075 standard; Protein; 1300 AA.
XX	ABB68075;
AC	ABB68075;
DT	26-MAR-2002 (first entry)
XX	Drosophila melanogaster polypeptide SEQ ID NO 31017.
DE	Drosophila; developmental biology; cell signalling; insecticide;
XX	pharmaceutical.
OS	Drosophila melanogaster.
XX	WO200171042-A2.
PN	27-SEP-2001.
XX	27-SEP-2001.
XX	23-MAR-2001; 2001WO-US09231.
PF	23-MAR-2000; 2000US-191637P.
XX	11-JUL-2000; 2000US-0614150.
PR	(PEKE) PE CORP NY.
XX	Venter JC, Adams M, Li PWD, Myers EW;
XX	WPI; 2001-656860/75.
XX	N-PSDB; ABL12178.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 31017; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1300 AA;
PS
Query Match 5.7%; Score 162.5; DB 22; Length 1300;
Best Local Similarity 20.6%; Pred. No. 0.0019;
Matches 99; Conservative 61; Mismatches 189; Indels 131; Gaps 18;
Qy 5 NLGHNPNVNSIIPAPPLPSQTDG---AGGRGQLINSTGPGSRALFTPVNSMADGDN 61
Db 438 NSTSNSNTNDSTGSETSTNGLVAGGAG-----GATGAAMLPTP---SQOSTGGK 488
Qy 62 RAS-----DVPGLPNPMLA-----ASEITLNDGFEVLHDHGPLDITLNRQ 102
Db 489 EATAVSVLLEKLPNVVVSPLTMKELRQKMTKYDAEIMANAAYQQQH----- 539
Qy 103 IGSSVFRVETQEDKHIAVQGRNGVETSVLSDOEYARLQSIDPEGKDKFVFTGGRGAG 162
Db 540 --QHHFHHHHHHHHHNGHGHASTGAEATAAQQQAAMQKPG-----VGGTGAAG 588
Qy 163 HAMVTASDITEARQRIILELLEPKGTGSGKAGSKGVGELRESGAENTTETOTSTST 222
Db 589 NAGATTVSSVA-----AGAGEVNGGRTSLURKMRVNS-----TSSSI 627
Qy 223 SSLRSDPKLMLALGTATGLIAGLAATGIVQALALTPEPDSPTTTDPAASATATARDQ 282
Db 628 STASADEVI-----APVVAASISLPSKAPVLMRCKPAQMAIALHQ 670
Qy 283 LTKEAFQNPDKQVNIDELGNAIPSGVLKDDVVAIBEQAKAAGEEAKQ-----A 333
Db 671 SQORQLRRSRQKEKLT-----GESSD---TSSEQOKK---EQKQDQHLQPKMFS 716
Qy 334 IENNAQAOKYDEQAQKEBELKVSSGAGYGLSGALILGGIGVAVTAALHKKNPVEQT 393
Db 717 LAEBEPQPKS--EEQOEQOKRVTRNSAGRVGL-----VARLATANNH-----IA 760
Qy 394 TTTTITTTTITGARTV---ENKPANNTPAQGNVDTPGSEDTMESRRSSMASTSTFFDTSS 450
Db 761 TTTSSSSSSSKATTITCNHNSNNSRNHNSNLSSLVSKRKPAPSEASSIPSSSTSS 820
RESULT 5
ABB66878
ID ABB66878 standard; Protein; 2586 AA.
XX
XX ABB66878;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 27426.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX

PN WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX
XX N-PSDB; ABL10981.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 27426; 21pp + Sequence Listing; English.
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 2586 AA;
SQ
Query Match 5.6%; Score 161; DB 22; Length 2586;
Best Local Similarity 19.8%; Pred. No. 0.0066;
Matches 99; Conservative 69; Mismatches 209; Indels 122; Gaps 17;
Qy 8 HNFVNNSIIPAPPLPSQTDGAGRGQLINSTGPGSRALFTPVNSMADGDNASDVP 67
Db 1123 HLPWVANPPHLPVVEVSQGTNGDNSSTOSS-----STTTTSSDEGQTTSSA- 1173
Qy 68 GLPNPMLAASEITLNDGFEVLHDHGPL-----DTLNRQIGSSVFRVETQEDGKH 118
Db 1174 --PVSTTTTTTTTTSSDGGQSTTSSDPVVEVSQGTNGDNSSTOSSSTTTTSSDGGQT 1231
Qy 119 IAV-----GQRNGVE-----TSVLSDOEYARLQSIDPEGKDKFVFTGGR 158
Db 1232 TSSAPVVEVTQSSSNGDGNSTQSTTTTTTTTTSSDGGESTTSSDPVVE---VSQGTN 1288
Qy 159 G-----GAGHAMVTASD---ITEARQRIILELLEPKGTGSGKAGSKGVGELRESNG 209
Db 1289 GDNSTOSSSTTTTTSSDGGQTTSSAPVVEVTQ-----GSSSNGDGNSTQSTTT 1339
Qy 210 AENTTETQSTSTSLRSDPKLMLALGTATGLIAGLAATGIVQALALTPEPDSPTTTD 269
Db 1340 TTTTITSSDGGESTTSSDPVVEVSQGT-----NGDNSST---Q 1375
Qy 270 AAASATATARDQLTKEAFQNPDKQVNIDELGNAIPSGVL-----KDDVV-----ANIEE 320
Db 1376 SSSSTTTTSSDGGQTTSSSGP-----VVDISQSSSNVLNDNDNNTNDIDFRWVANPPP 1429
Qy 321 QAKAAGEEAKQAENNAQAOKYDEQAQKEBELKVSSGA-----GYGLSGALLGGGI 375
Db 1430 YLTPLWKSVEQMAITAPLSLRPPQOQTSSDEGQTTSSAPVVDISQSS-----SNGD 1484
Qy 376 GVAVTAALHKKNPVEQT-----TTTTTTTTSARTVKNKPNANTPAQGNVDTPGSEDTMESRR 435
Db 1485 GNST-----QSSTTTTTTTTTSSDGGESTTILSDPVVEVSQGTNGDNSSTQSS 1532
Qy 436 SSMASSTSTFPDTSIGGP 454

PR	08-OCT-1998;	98US-0103633.	CC	polypeptides from the present invention.
PR	08-OCT-1998;	98US-0103678.	XX	
PR	08-OCT-1998;	98US-0103679.	SQ	Sequence 596 AA;
PR	08-OCT-1998;	98US-0103711.		
PR	14-OCT-1998;	98US-0104257.		
PR	20-OCT-1998;	98US-0104987.		
PR	20-OCT-1998;	98US-0105000.		
PR	20-OCT-1998;	98US-0105002.		
PR	21-OCT-1998;	98US-0105104.		
PR	22-OCT-1998;	98US-0105169.		
PR	22-OCT-1998;	98US-0105266.		
PR	22-OCT-1998;	98US-0105693.		
PR	26-OCT-1998;	98US-0105694.		
PR	27-OCT-1998;	98US-0105807.		
PR	27-OCT-1998;	98US-0105881.		
PR	27-OCT-1998;	98US-0105882.		
PR	27-OCT-1998;	98US-0106033.		
PR	28-OCT-1998;	98US-0106023.		
PR	28-OCT-1998;	98US-0106024.		
PR	28-OCT-1998;	98US-0106029.		
PR	28-OCT-1998;	98US-0106030.		
PR	28-OCT-1998;	98US-0106032.		
PR	28-OCT-1998;	98US-0106033.		
PR	28-OCT-1998;	98US-0106178.		
PR	29-OCT-1998;	98US-0106248.		
PR	29-OCT-1998;	98US-0106384.		
PR	29-OCT-1998;	98US-0108500.		
PR	30-OCT-1998;	98US-0106464.		
PR	03-NOV-1998;	98US-0106856.		
PR	03-NOV-1998;	98US-0106902.		
PR	03-NOV-1998;	98US-0106905.		
PR	03-NOV-1998;	98US-0106919.		
PR	03-NOV-1998;	98US-0106932.		
PR	10-NOV-1998;	98US-0106934.		
PR	10-NOV-1998;	98US-0107783.		
PR	17-NOV-1998;	98US-0108775.		
PR	17-NOV-1998;	98US-0108779.		
PR	17-NOV-1998;	98US-0108787.		
PR	17-NOV-1998;	98US-0108788.		
PR	17-NOV-1998;	98US-0108801.		
PR	17-NOV-1998;	98US-0108802.		
PR	17-NOV-1998;	98US-0108806.		
PR	17-NOV-1998;	98US-0108807.		
PR	17-NOV-1998;	98US-0108867.		
PR	17-NOV-1998;	98US-0108925.		
PR	18-NOV-1998;	98US-0108848.		
PR	18-NOV-1998;	98US-0108849.		
PR	18-NOV-1998;	98US-0108850.		
PR	18-NOV-1998;	98US-0108851.		
PR	18-NOV-1998;	98US-0108852.		
PR	18-NOV-1998;	98US-0108858.		
PR	18-NOV-1998;	98US-0108904.		
XX				
FA	(GETH) GENENTECH INC.			
XX				
PI	Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WI;			
XX				
XX	WPI; 2000-237871/20.			
DR	N-PSDB; AAA37090.			
DR				
XX				
PT	New mammalian DNA sequences encoding transmembrane, receptor or			
PT	secreted PRO polypeptides, useful for screening of potential peptide or			
PT	small molecule inhibitors of the relevant receptor/ligand interactions			
XX				
PS	Claim 12; Fig 138; 773pp; English.			
XX				
CC	AAA37022 to AAA37144 encode the new isolated human transmembrane,			
CC	receptor or secreted PRO polypeptides given in AAY99340 to AAY99462. The			
CC	transmembrane and receptor PRO proteins can be used for screening of			
CC	potential peptide or small molecule inhibitors of the relevant			
CC	receptor/ligand interactions. The polypeptides and nucleotide sequences			
CC	encoding then have various industrial applications, including uses as			
CC	pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent			
CC	PCR primers and hybridisation probes used in the isolation of the PRO			

Qy	4	GNLGNPNVANSIPPAPPLPSQTDGAGGRQQLINSTGLGSRALFTVRNSMADSGNRA	63
Db	13	GLLLHLEAATNS-----NETSTANTGSSVSSG-----ASTATNSG	49
Qy	64	SDVPEGLPVNPMRLAASEITLNDGFEVLHDHGPDLTLNRQIGSSV-FRVEQEDGKHIAVG	122
Db	50	SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEPHITSSGISTATNSEFSSTASSGISATN	107
Qy	123	QRNGVETS--VLSDOEYARLQSIDPEGDKFVFTGGRGGAGHAMVTVASDITEARORIL	180
Db	108	SESTTSSGASTATNSE-----SSTPSSGASTVTVNSGSSVTSSGASTATNSESTVSSRAS	163
Qy	181	ELLEPKGTGESKGAGESKGVGELRESNSGAENTTETOTSTSTSLRSDPKLMLALGTAT	240
Db	164	TATNSESTLSSGASTATN-SDSSTTSSGASTATNSESTTSSG-----ASTAT	211
Qy	241	GLIGLAATGIVQALATPEPDPSTTTDDAAASATETATRDQLTKBAFQNPDKQKNIDE	300
Db	212	N-----SESSTVSSKRATATNSESTT---SSGASTATNSESTTSSGAGTATNSESTTS	264
Qy	301	LG-----NAIPSGVLKDDVVANIEEQAKAAGEAKQAIENNAQAQKYDEQQAQR	351
Db	265	SGASTATNSDSSTVSSGA---STATNSESTTSSGAST-----ATN	302
Qy	352	QEELKVSSSGAGYGL-SCALILGGIGVAVTAALHRKNQPVQETTTTTTTTTTSARTVEN	410
Db	303	SESTTSSGASTATNSDSSTTSSGAGTATNSESTVSSGISTVTVNSSESTPSSGANTATN	362
Qy	411	KPANTPAQGNVDTPGSEDTMESRRSSMASTSTFFDTSSIGGFCRIRMLMLKRCMIR	470
Db	363	SESTTSSGANTATNSESTVSSGASTATNSESS---TTSSG-----	401
Qy	471	CRLLILRLPRIWGIQISVVYSTIOHPPTDTTNGARLLGNPSAGIOSTVARLA-----	524
Db	402	-----VSTATNSESTTSSGASTATNSDSSTTSSGASTATNSESS	441
Qy	525	-LSGGL-----RHDWGGTLTGGNSAVNTSNPPAPGS	555
Db	442	TVSSGISTVTVNSSESTTSSGANTATNSGSSVTSAGS	477

RESULT 9	
AAU29178	
ID	AAU29178 standard; Protein; 596 AA.
XX	
AC	AAU29178;
XX	
DT	18-DEC-2001 (first entry)
XX	
DE	Human PRO polypeptide sequence #155.
XX	
KW	PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
KW	dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
KW	blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
KW	adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
OS	Homo sapiens.
XX	
PN	WO200168848-A2.
XX	
PD	20-SEP-2001.
XX	
PF	28-FEB-2001; 2001WO-US06520.
XX	
PR	01-MAR-2000; 2000WO-US05601.

PR 02-MAR-2000; 2000WO-US05841.
 PR 03-MAR-2000; 2000US-187202P.
 PR 06-MAR-2000; 2000US-186968P.
 PR 14-MAR-2000; 2000US-189320P.
 PR 15-MAR-2000; 2000US-189328P.
 PR 15-MAR-2000; 2000WO-US06884.
 PR 21-MAR-2000; 2000US-190828P.
 PR 21-MAR-2000; 2000US-191007P.
 PR 21-MAR-2000; 2000US-191048P.
 PR 21-MAR-2000; 2000US-191314P.
 PR 28-MAR-2000; 2000US-192655P.
 PR 29-MAR-2000; 2000US-193032P.
 PR 29-MAR-2000; 2000US-193053P.
 PR 30-MAR-2000; 2000WO-US08439.
 PR 04-APR-2000; 2000US-194443P.
 PR 04-APR-2000; 2000US-194647P.
 PR 11-APR-2000; 2000US-195975P.
 PR 11-APR-2000; 2000US-196000P.
 PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 PA (GETH) GENENTECH INC.
 XX
 XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 DR WPI; 2001-602746/68.
 DR N-PSDB; AAS46079.
 XX
 PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 PS Claim 11; Fig 310; 774pp; English.
 XX
 CC Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
 CC The PRO polypeptides and their associated nucleic acids can be used to
 CC detect the presence of a tumour in a mammal by comparing the level of
 CC expression of a PRO polypeptide in a test sample of cells from the animal
 CC and a control sample of normal cells, whereby a higher level of
 CC expression in the test sample indicates the presence of a tumour in the
 CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
 CC and rabbits but are preferably human. The polypeptides can be used to
 CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
 CC when contacted with it. A specific polypeptide can be used to stimulate
 CC the proliferation or differentiation of chondrocyte cells. The PRO
 CC proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 XX
 SQ Sequence 596 AA;
 XX
 Query Match 5.2%; Score 148.5; DB 22; Length 596;
 XX

Best Local Similarity 17.7%; Pred. No. 0.0081;
 Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;
 QY 4 GNLGHNPNVNNSTPPAPPLPSOTDGGAGRGGLNSTGPLGSRALFTPVNNSMADSGDNRA 63
 DB 13 GLLHLEAATNS-----NETSTANTGSSVSSG-----ASTATNSG 49
 QY 64 SDVPGLPVNPMLAASEITLNDGFEVLHDHGPDLDTLNRQIGSSV-FRVTFQEDGKHIAVG 122
 DB 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEPHITSSGISTATNSEFSTASSGIGSIATN 107
 QY 123 QRNGVETS--VVLSDOEYARLQSIDPEGKDKFVFTGGRGAGHAMVTVASDIETEARQRI 180
 DB 108 SESSTTSSGASTATNSE---SSTPSSGASTVTNCGSSVTSSTSGASTATNSESTVSSRAS 163
 QY 181 ELLEPKGTGSKGAGESKVGELRESNGAENTTQTSTSTSSLSRSDPKLWLALGTAT 240
 DB 164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESTTSSG-----ASTAT 211
 QY 241 GLIGLAATGIVQALALTPEPDSPTTTPDPAASATETATRDQLTKEAFQNPDKQKNIDE 300
 DB 212 N-----SESSTVSSRASTATNSESTT---SSGASTATNSESTTSSGASTATNSESTTS 264
 QY 301 LG-----NAIPSGVLKDDVVANIEBOAKAAGEAKQQAENNAQAOKKYDEQOAKR 351
 DB 265 SGASTATNSDSTVSSGA---STATNSESTTSSGAST-----ATN 302
 QY 352 QBELKVSSGAGYGL-SCALLILGGIGVAVTAALHRKNQPVQTTTTTTTTTTTSARTVEN 410
 DB 303 SESSTTSSGASTATNSDSTTSSGAGTATNSESTVSSGISTVTNSESSTPSSGANTATN 362
 QY 411 KPNATPAQGNVDTPGSEDTMESRRSSMASTSTFFDTSSIGGPCRIRMLMLKHCMMIR 470
 DB 363 SESSTTSSGANTATNSESTVSSGASTATNSESS---TTSSG----- 401
 QY 471 CRLLILRLPRWIGIQLSVVYTIQHPDRTTNGARLLGNPAGIOSTVARLA----- 524
 DB 402 -----VSTATNSESTTSSGASTATNSDSTTSSGASTATNSESTTSSGASTATNSESS 441
 QY 525 -LSGGL-----RHDMMGLTGGSNSAVNTSNPPAPGS 555
 DB 442 TVSSGISTVTNSESSTTSSGANTATNSGSSVTSAGS 477
 RESULT 10
 AAB87575
 ID AAB87575 standard; Protein; 596 AA.
 XX
 AC AAB87575;
 XX
 DT 15-MAY-2001 (first entry)
 XX
 DE Human PRO1342.
 XX
 KW Human; PRO protein; mapping.
 XX
 OS Homo sapiens.
 XX
 PN WO2000116318-A2.
 XX
 PD 08-MAR-2001.
 XX
 PF 24-AUG-2000; 2000WO-US23328.
 XX
 PR 01-SEP-1999; 99WO-US20111.
 PR 15-SEP-1999; 99WO-US21090.
 PR 07-DEC-1999; 99US-0169495.
 PR 09-DEC-1999; 99US-0170262.
 PR 11-JAN-2000; 2000US-0175481.
 PR 18-FEB-2000; 2000WO-US04341.
 PR 18-FEB-2000; 2000WO-US04342.
 PR 22-FEB-2000; 2000WO-US04414.
 PR 01-MAR-2000; 2000WO-US05601.


```
PR 03-MAR-2000; 2000US-0187202.
PR 25-APR-2000; 2000US-0199397.
PR 22-MAY-2000; 2000WO-US14042.
PR 05-JUN-2000; 2000US-0209832.
XX
XX
XX (GETH ) GENENTECH INC.
XX
XX Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
XX Grimaldi CJ, Gurney AL, Watanabe CK, Wood WI;
XX N-PSDB; AAF92107.
XX
XX WPI; 2001-183260/18.
XX
XX Eighty four nucleic acids encoding PRO polypeptides, useful in
XX molecular biology, including use as hybridization probes, and in
XX chromosome and gene mapping.
XX
XX Claim 12; Fig 100; 278pp; English.
XX
XX The present sequence is a human PRO polypeptide (secreted and
XX transmembrane). The PRO protein, and PRO agonists, PRO antagonists or
XX anti-PRO antibodies are useful for preparation of a medicament useful in
XX the treatment of a condition which is responsive to the PRO protein.
XX agonists, antagonists or anti-PRO antibodies. The PRO protein may also be
XX employed as molecular weight markers for protein electrophoresis. The PRO
XX coding sequence has applications in molecular biology, including use as
XX hybridisation probes, and in chromosome and gene mapping.
XX
XX Sequence 596 AA;
XX
XX Query Match 5.2%; Score 148.5; DB 22; Length 596;
XX Best Local Similarity 17.7%; Pred. No. 0.0081;
XX Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;
XX
XX 4 GNIGHNPNVNSIPPAPPLPSQTDGAGRGQLNSTGPGSLRALFTPVNSMADSGNRA 63
XX 13 GLLHLEAATNS-----NETSTANTSSVSSG-----ASTATNSG 49
XX
XX 64 SDVPGPLVPMRLAASEITLNDGFEVLHDHGLDPLTLNRQIGSSV-FRVETQEDGKHIAG 122
XX 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSFEFTASSGISIATN 107
XX
XX 123 QRNGVETS--VVLSDOEYARLQSIDPEGKDKFVFTGGRGAGHAMVTVASDITEARQIL 180
XX 108 SESSTSSGASTATNSE---SSTPSSGASTVTVNSGSSVTSSGASTATNSSESTVSSRAS 163
XX
XX 181 ELLEPKTGCKSGKSGVGEELRESNGAENTTETSTSTSLRSDPKLWLALGTAVT 240
XX 164 TATNSSESTLSSGASTATN-SDSSTTSSGASTATNSSESTSSG-----ASTAT 211
XX
XX 241 GLIGLAATGIVQALALTPEDSPPTTDPDAAASATETATRDQLTKEAFQNPDKVNI 300
XX 212 N-----SESTVSSRASTATNSESTT---SGGASTATNSRRTSNGAGTATNSESTTS 264
XX
XX 301 LG-----NAITPSGLVKDDVVANIEEQAAGEAKQAIAENNAQAOKKYDEQOAKR 351
XX 265 SGASTATNSDSTVSSGA---STATNSSESTSSGAST-----ATN 302
XX
XX 352 QBEELKVSAGAGYGL-SCALILGGIGVAVTAALHRKNQVPEQITTTTTTTTSARIVEN 410
XX 303 SESSTSSGASTATNSDSTSSGASTATNSSESTVSSGISTVTVNSSESTPSSGANTATN 362
XX
XX 411 KPAANTPAQGNVDTPGSEDTMESRRSSMAWTSSTFFDPTSSIGGPCRIRLMLKHCIRM 470
XX 363 SESSTSSGANTATNSSESTVSSGASTATNSESS---TSSG----- 401
XX
XX 471 CRLILIRLFRWIGIQISVYVSTTQHPPRDITDNGARLLGNPSAGIQSTYARLA----- 524
XX 402 -----VSTATNSSESTTSSGASTATNSDSTSSGASTATNSESS 441
XX
XX 525 -LSGGL-----RHDMGGLTGGSNSAVNTSNPPAPGS 555
XX 442 TVSSGISTVTVNSSESTTSSGANTATNSGSSVTSAGS 477
```

RESULT 11

```
AAB66157
XX
XX AAB66157 standard; protein; 596 AA.
XX
XX AC AAB66157;
XX
XX DT 02-APR-2001 (first entry)
XX
XX DE Protein of the invention #69.
XX
XX KW Secreted; transmembrane; gene therapy.
XX
XX OS Unidentified.
XX
XX PN WO200078961-A1.
XX
XX PD 28-DEC-2000.
XX
XX 18-FEB-2000; 2000WO-US04342.
XX
XX 23-JUN-1999; 99US-0141037.
XX 20-JUL-1999; 99US-0144758.
XX 26-JUL-1999; 99US-0145698.
XX 01-SEP-1999; 99WO-US20111.
XX 23-OCT-1999; 99US-0162506.
XX 30-NOV-1999; 99WO-US28313.
XX 02-DEC-1999; 99WO-US28551.
XX 16-DEC-1999; 99WO-US30095.
XX 03-JAN-2000; 2000WO-US00219.
XX 06-JAN-2000; 2000WO-US00376.
XX
XX (GETH ) GENENTECH INC.
XX
XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;
XX Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;
XX Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;
XX Watanabe CK, Williams PM, Wood WI;
XX
XX WPI; 2001-071395/08.
XX
XX Secreted and transmembrane proteins and nucleic acids designated PRO,
XX useful as hybridization probes, in chromosome and gene mapping and gene
XX therapy -
XX
XX Claim 1; Fig 138; 787pp; English.
XX
XX The present invention relates to secreted and transmembrane proteins.
XX These proteins and the DNA encoding them may be used as hybridization
XX probes, in chromosome and gene mapping and in the generation of
XX anti-sense RNA and DNA. They may also be used to generate either
XX transgenic animals or knockout animals which are in turn useful for
XX diagnostic and screening of therapeutically useful reagents.
XX The nucleic acids may also be used in gene therapy.
XX
XX Sequence 596 AA;
XX
XX Query Match 5.2%; Score 148.5; DB 22; Length 596;
XX Best Local Similarity 17.7%; Pred. No. 0.0081;
XX Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;
XX
XX 4 GNIGHNPNVNSIPPAPPLPSQTDGAGRGQLNSTGPGSLRALFTPVNSMADSGNRA 63
XX 13 GLLHLEAATNS-----NETSTANTSSVSSG-----ASTATNSG 49
XX
XX 64 SDVPGPLVPMRLAASEITLNDGFEVLHDHGLDPLTLNRQIGSSV-FRVETQEDGKHIAG 122
XX 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEFHTTSSGISTATNSFEFTASSGISIATN 107
XX
XX 123 QRNGVETS--VVLSDOEYARLQSIDPEGKDKFVFTGGRGAGHAMVTVASDITEARQIL 180
XX 108 SESSTSSGASTATNSE---SSTPSSGASTVTVNSGSSVTSSGASTATNSSESTVSSRAS 163
XX
XX 181 ELLEPKTGCKSGKSGVGEELRESNGAENTTETSTSTSLRSDPKLWLALGTAVT 240
XX 164 TATNSSESTLSSGASTATN-SDSSTTSSGASTATNSSESTSSG-----ASTAT 211
XX
XX 241 GLIGLAATGIVQALALTPEDSPPTTDPDAAASATETATRDQLTKEAFQNPDKVNI 300
XX 212 N-----SESTVSSRASTATNSESTT---SGGASTATNSRRTSNGAGTATNSESTTS 264
XX
XX 301 LG-----NAITPSGLVKDDVVANIEEQAAGEAKQAIAENNAQAOKKYDEQOAKR 351
XX 265 SGASTATNSDSTVSSGA---STATNSSESTSSGAST-----ATN 302
XX
XX 352 QBEELKVSAGAGYGL-SCALILGGIGVAVTAALHRKNQVPEQITTTTTTTTSARIVEN 410
XX 303 SESSTSSGASTATNSDSTSSGASTATNSSESTVSSGISTVTVNSSESTPSSGANTATN 362
XX
XX 411 KPAANTPAQGNVDTPGSEDTMESRRSSMAWTSSTFFDPTSSIGGPCRIRLMLKHCIRM 470
XX 363 SESSTSSGANTATNSSESTVSSGASTATNSESS---TSSG----- 401
XX
XX 471 CRLILIRLFRWIGIQISVYVSTTQHPPRDITDNGARLLGNPSAGIQSTYARLA----- 524
XX 402 -----VSTATNSSESTTSSGASTATNSDSTSSGASTATNSESS 441
XX
XX 525 -LSGGL-----RHDMGGLTGGSNSAVNTSNPPAPGS 555
XX 442 TVSSGISTVTVNSSESTTSSGANTATNSGSSVTSAGS 477
```


Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;	
Qy 4	GNLGNPNVNNISIPPAPLPQSDGAGRGQLINSTGLGSRALFTFVRNSMADSGNRA 63
Db	
Db 13	GLLLHLEAATNS-----NETSANTGSSVISSG-----ASTATNSG 49
Qy 64	SDVPLPVPNRLAASEITLNDGFEVLHDGPLDTLNRQIGSSV-FRVEIQEDGKHIAVG 122
Db	
Db 50	SSVTSSGVSTATIGSSVTSN-GVSIIV-TNSEFTTSSGISTATNSEFSTASSGISIATN 107
Qy 123	QRNGVETS--VVLSDOEVARLQSDPEKGRKFVTGGRGGAGHAMVTVASDITEAROKIL 180
Db	
Db 108	SESTTSGASTATNSE-----SSTPSSGASTVTVNGSSVTVSSGASTATNSESTVSSRAS 163
Qy 181	ELLEPKGTGSGKAGESKVGELRESNGBAENTTETOTSTSTSLRSDPKLWLALGTAVT 240
Db	
Db 164	TATNSESTLSSGASTATN-SDSSTTSSGASTATNSESTTSSG-----ASTAT 211
Qy 241	GLIGLATGIQVALALPEPDSPTTTPDAAASATETATRDQLTKEAFQNPDPNOKVNI 300
Db	
Db 212	N-----SESSTVSSRASTATNSESTT---SSGASTATNSESTTSSGAGTATNSESTTS 264
Qy 301	LG-----NAIPSGVLKDDVNVANIEEQAKAAGEAKQQAIEENNAQAQKDYDEQAKR 351
Db	
Db 265	SGASTATNSDSTVSSGA---STATNSESTTSSGAST-----ATN 302
Qy 352	QEELKVSAGGAGYGL-SGAILGGIGIGVAVTAALHRKNQPVQTTTITTTTTTSARTVEN 410
Db	
Db 303	SESSTSSGASTATNSDSTTSSGAGTATNSESTVSSGISTVTVNSESTPSSGANTATN 362
Qy 411	KPANNTPAQGNVDTPGSEDNMESSRMASTSTFTFTSTTSGGPCIRIMLMKHCWIRR 470
Db	
Db 363	SESSTSSGASTATNSESTVSSGASTATNSESS---TSSG----- 401
Qy 471	CRLLILRLFRWIGIQISVTVSTIQHPRTDTDNGARLLGNPSAGIQSTVARLA----- 524
Db	
Db 402	-----VSTATNSESTTSSGASTATNSDSTTSSSEASTATNSESS 441
Qy 525	-LSGGL-----RHDGMGLTGGSNAVTNSNPPAPGS 555
Db	
Db 442	TVSSGISTVTVNSESTTSSGANTATNSGSSVTSAGS 477
RESULT 13	
ID	ABU71266
XX	ABU71266 standard; Protein; 596 AA.
AC	ABU71266;
XX	
DT	10-JUN-2003 (first entry)
XX	
DE	Human PRO1342 protein.
XX	
KW	Human; PRO; secreted; transmembrane; cytotstatic; TNF-alpha; blood;
KW	tumour necrosis factor alpha release; chondrocyte cell; proliferation;
KW	differentiation; tumour; gene therapy.
XX	
OS	Homo sapiens.
XX	
FN	US2003036143-A1.
XX	
PD	20-FEB-2003.
XX	
PF	02-JUL-2002; 2002US-0187600.
XX	
PR	16-SEP-1998; 98WO-US19330.
PR	07-OCT-1998; 98WO-US21141.
PR	01-DEC-1998; 98WO-US25108.
PR	08-MAR-1999; 99WO-US05028.
PR	14-MAY-1999; 99WO-US10733.
PR	02-JUN-1999; 99WO-US12252.
PR	01-SEP-1999; 99WO-US20111.
PR	15-SEP-1999; 99WO-US21090.

PR 01-DEC-1999;	99WO-US28301.
PR 02-DEC-1999;	99WO-US28551.
PR 30-DEC-1999;	99WO-US31274.
PR 05-JAN-2000;	2000WO-US00219.
PR 18-FEB-2000;	2000WO-US04341.
PR 22-FEB-2000;	2000WO-US04342.
PR 28-FEB-2000;	2000WO-US04414.
PR 24-FEB-2000;	2000WO-US05004.
PR 01-MAR-2000;	2000WO-US05601.
PR 02-MAR-2000;	2000WO-US05841.
PR 15-MAR-2000;	2000WO-US06884.
PR 30-MAR-2000;	2000WO-US06843.
PR 17-MAY-2000;	2000WO-US13705.
PR 22-MAY-2000;	2000WO-US14042.
PR 30-MAY-2000;	2000WO-US14941.
PR 02-JUN-2000;	2000WO-US15264.
PR 28-JUL-2000;	2000WO-US20710.
PR 24-AUG-2000;	2000WO-US23328.
PR 08-NOV-2000;	2000WO-US30952.
PR 01-DEC-2000;	2000WO-US32678.
PR 20-DEC-2000;	2000WO-US34956.
PR 28-FEB-2001;	2001WO-US06520.
PR 01-JUN-2001;	2001WO-US17800.
PR 20-JUN-2001;	2001WO-US19692.
PR 29-JUN-2001;	2001WO-US21066.
PR 09-JUL-2001;	2001WO-US21735.
PR 29-AUG-2001;	2001WO-US27099.
PR 18-SEP-1997;	97US-059263P.
PR 18-SEP-1997;	97US-059266P.
PR 17-OCT-1997;	97US-062250P.
PR 21-OCT-1997;	97US-063486P.
PR 24-OCT-1997;	97US-063120P.
PR 24-OCT-1997;	97US-063121P.
PR 28-OCT-1997;	97US-063540P.
PR 28-OCT-1997;	97US-063541P.
PR 28-OCT-1997;	97US-063544P.
PR 28-OCT-1997;	97US-063564P.
PR 31-OCT-1997;	97US-063734P.
PR 31-OCT-1997;	97US-063870P.
PR 31-OCT-1997;	97US-064103P.
PR 13-NOV-1997;	97US-065311P.
PR 21-NOV-1997;	97US-066120P.
PR 24-NOV-1997;	97US-066466P.
PR 11-DEC-1997;	97US-066772P.
PR 12-DEC-1997;	97US-069335P.
PR 17-DEC-1997;	97US-069425P.
PR 18-DEC-1997;	97US-069870P.
PR 10-MAR-1998;	98US-068017P.
PR 11-MAR-1998;	98US-077632P.
PR 11-MAR-1998;	98US-077649P.
PR 20-MAR-1998;	98US-078886P.
PR 20-MAR-1998;	98US-078939P.
PR 27-MAR-1998;	98US-079664P.
PR 27-MAR-1998;	98US-079786P.
PR 31-MAR-1998;	98US-080107P.
PR 31-MAR-1998;	98US-080194P.
PR 01-APR-1998;	98US-080327P.
PR 01-APR-1998;	98US-080333P.
PR 08-APR-1998;	98US-081049P.
PR 08-APR-1998;	98US-081070P.
PR 09-APR-1998;	98US-081195P.
PR 15-APR-1998;	98US-081838P.
PR 21-APR-1998;	98US-082568P.
PR 21-APR-1998;	98US-082569P.
PR 22-APR-1998;	98US-082704P.
PR 22-APR-1998;	98US-082797P.
PR 28-APR-1998;	98US-083322P.
PR 29-APR-1998;	98US-083495P.
PR 29-APR-1998;	98US-083496P.
PR 29-APR-1998;	98US-083499P.
PR 05-MAY-1998;	98US-083559P.
PR	98US-084366P.

PR	06-MAY-1998;	98US-084414P.	PR	10-AUG-1998;	98US-095998P.
PR	07-MAY-1998;	98US-084639P.	PR	10-AUG-1998;	98US-096012P.
PR	07-MAY-1998;	98US-084640P.	PR	17-AUG-1998;	98US-096757P.
PR	07-MAY-1998;	98US-084643P.	PR	17-AUG-1998;	98US-096766P.
PR	15-MAY-1998;	98US-085573P.	PR	17-AUG-1998;	98US-096867P.
PR	15-MAY-1998;	98US-085580P.	PR	17-AUG-1998;	98US-096891P.
PR	15-MAY-1998;	98US-085582P.	PR	18-AUG-1998;	98US-096897P.
PR	15-MAY-1998;	98US-085700P.	PR	18-AUG-1998;	98US-096949P.
PR	18-MAY-1998;	98US-086023P.	PR	18-AUG-1998;	98US-096959P.
PR	22-MAY-1998;	98US-086392P.	PR	18-AUG-1998;	98US-097022P.
PR	22-MAY-1998;	98US-086486P.	PR	26-AUG-1998;	98US-097552P.
PR	28-MAY-1998;	98US-087098P.	PR	26-AUG-1998;	98US-097555P.
PR	28-MAY-1998;	98US-087208P.	PR	26-AUG-1998;	98US-097971P.
PR	02-JUN-1998;	98US-087609P.	PR	26-AUG-1998;	98US-097974P.
PR	02-JUN-1998;	98US-087759P.	PR	26-AUG-1998;	98US-098014P.
PR	03-JUN-1998;	98US-087827P.	PR	26-AUG-1998;	98US-098016P.
PR	04-JUN-1998;	98US-088025P.	PR	01-SEP-1998;	98US-098716P.
PR	04-JUN-1998;	98US-088028P.	PR	01-SEP-1998;	98US-098723P.
PR	04-JUN-1998;	98US-088029P.	PR	02-SEP-1998;	98US-098803P.
PR	04-JUN-1998;	98US-088033P.	PR	02-SEP-1998;	98US-098821P.
PR	04-JUN-1998;	98US-088326P.	PR	02-SEP-1998;	98US-098843P.
PR	05-JUN-1998;	98US-088167P.	PR	09-SEP-1998;	98US-099602P.
PR	05-JUN-1998;	98US-088202P.	PR	10-SEP-1998;	98US-099741P.
PR	05-JUN-1998;	98US-088212P.	PR	10-SEP-1998;	98US-099754P.
PR	05-JUN-1998;	98US-088217P.	PR	10-SEP-1998;	98US-099763P.
PR	09-JUN-1998;	98US-088655P.	PR	10-SEP-1998;	98US-099812P.
PR	10-JUN-1998;	98US-088722P.			
PR	10-JUN-1998;	98US-088738P.			
PR	10-JUN-1998;	98US-088740P.			
PR	10-JUN-1998;	98US-088811P.			
PR	10-JUN-1998;	98US-088824P.			
PR	10-JUN-1998;	98US-088825P.			
PR	10-JUN-1998;	98US-088826P.			
PR	11-JUN-1998;	98US-088861P.			
PR	11-JUN-1998;	98US-088863P.			
PR	11-JUN-1998;	98US-088876P.			
PR	12-JUN-1998;	98US-089090P.			
PR	12-JUN-1998;	98US-089105P.			
PR	16-JUN-1998;	98US-089512P.			
PR	16-JUN-1998;	98US-089514P.			
PR	17-JUN-1998;	98US-089538P.			
PR	17-JUN-1998;	98US-089598P.			
PR	17-JUN-1998;	98US-089653P.			
PR	18-JUN-1998;	98US-089908P.			
PR	19-JUN-1998;	98US-089952P.			
PR	22-JUN-1998;	98US-090246P.			
PR	22-JUN-1998;	98US-090252P.			
PR	22-JUN-1998;	98US-090254P.			
PR	24-JUN-1998;	98US-090429P.			
PR	24-JUN-1998;	98US-090435P.			
PR	24-JUN-1998;	98US-090444P.			
PR	24-JUN-1998;	98US-090461P.			
PR	24-JUN-1998;	98US-090535P.			
PR	24-JUN-1998;	98US-090540P.			
PR	25-JUN-1998;	98US-090676P.			
PR	25-JUN-1998;	98US-090678P.			
PR	25-JUN-1998;	98US-090688P.			
PR	25-JUN-1998;	98US-090690P.			
PR	25-JUN-1998;	98US-090694P.			
PR	25-JUN-1998;	98US-090695P.			
PR	25-JUN-1998;	98US-090696P.			
PR	26-JUN-1998;	98US-090862P.			
PR	26-JUN-1998;	98US-090863P.			
PR	26-JUN-1998;	98US-091010P.			
PR	01-JUL-1998;	98US-091359P.			
PR	01-JUL-1998;	98US-091544P.			
PR	02-JUL-1998;	98US-091478P.			
PR	02-JUL-1998;	98US-091486P.			
PR	02-JUL-1998;	98US-091626P.			
PR	02-JUL-1998;	98US-091628P.			
PR	02-JUL-1998;	98US-091632P.			
PR	24-JUL-1998;	98US-094006P.			
PR	04-AUG-1998;	98US-095282P.			

Qy	4	GNLGNPNVNNISPPAPLPSTQDAGGROQLINSTGPLGRALFTFVRNMSADSGNRA	63
Db	13	GLLLHLEATNS-----NETSTANTGSSVSSG-----ASTATNSG	49
Qy	64	SDVPGLPWNRLAASEITLNDGFEVLHDGCPDLTLNRQIGSSV-FRVETQEDGKHAVG	122
Db	50	SSVTSGCVSTATISGSSVTSN-GVSIV-TNSEPHITSSGISTATNSEFSTASSGISATN	107
Qy	123	QRNGVETS--VLSDOYARLQSIDPEGKDFVFTGGRGAGHAMVTVASDIPEARORIL	180
Db	108	SESTSSGASTATNSE---SSTPSSGASTVTVNSGSSVTSSGASTATNSESTVSSRAS	163
Qy	181	ELLEPKCTGESKAGESKGVGELRENSGAENTTETOTSTSTSLRSDPKLMLALGTVAT	240
Db	164	TATNSESTLSSGASTATN-SDSSTTSSGASTATNSESTSSG-----ASTAT	211
Qy	241	GLIGLAATGIVQALALTPEPDSPTTTPDAAASATETATRDQLTKEAFQNPDKQKNIDE	300
Db	212	N---SESTVSRASTATNSESTT---SSGASTATNSESTRTNSGAGTATNSESTTS	264
Qy	301	LG-----NAIPSGVLKDDVVANIEEQAKAEEAKQQAENNAQAQKQYDEQQAQR	351
Db	265	SGASTATNSDSSTVSSGA---STATNSESTSSGAST-----ATN	302
Qy	352	QEELKVSSGAGYGL-SCALILGGIGVAVTAALHRKNQPVQTTTTTTTTTTSARTVEN	410
Db	303	SESTTSSGASTATNSDSSTTSSGAGTATNSESTVSSGISTVTVNSESTPSSGANTATN	362
Qy	411	KPANTPAQGNVDTPGSEDTMESRRSSMASTSTFFDTSSIGGPCRLMLMLKRCMIRR	470
Db	363	SESTTSSGANTATNSESTVSSGASTATNSESS---TTSSG-----	401
Qy	471	CRLLILRLFRWGIQISVVVYTIQHPPTDNGARLLGNPSAGIOSTVARLA-----	524
Db	402	-----VSTATNSESTTSSGASTATNSDSSTTSSSEASTATNSESS	441
Qy	525	-LSCGL-----RHDMGGLTGGNSNAVNTNNPPAPGS	555
Db	442	TVSSGISTVTVNSESTTSSGANTATNSGSSVTSAGS	477

ABU71555
ID ABU71555 standard; Protein; 596 AA.
XX AC
XX ABU71555;
XX DT
XX 10-JUN-2003 (first entry)
XX DE Human secreted polypeptide PRO1342.
XX KW Human; gene therapy; tumour; cancer.
XX OS Homo sapiens.
XX PN US2003013855-A1.
XX PD 16-JAN-2003.
XX PF 03-MAY-2002; 2002US-00631616.
XX PR 30-DEC-1998; 98KR-00621142.
XX PR 08-MAR-1999; 99WO-US05028.
XX PR 14-MAY-1999; 99WO-US10733.
XX PR 30-DEC-1999; 99WO-US11273.
XX PR 18-FEB-2000; 2000WO-US04341.
XX PR 01-MAR-2000; 2000WO-US05601.
XX PR 02-MAR-2000; 2000WO-US05841.
XX PR 21-MAR-2000; 2000WO-US07532.
XX PR 22-MAY-2000; 2000WO-US14042.
XX PR 02-JUN-2000; 2000WO-US15264.
XX PR 24-AUG-2000; 2000WO-US23328.
XX PR 10-NOV-2000; 2000WO-US30873.
XX PR 01-DEC-2000; 2000WO-US32678.
XX PR 20-DEC-2000; 2000WO-US34956.
XX PR 28-FEB-2001; 2001WO-US06520.
XX PR 01-JUN-2001; 2001WO-US17800.
XX PR 14-MAY-1999; 99US-0311832.
XX PR 25-AUG-1999; 99US-0380137.
XX PR 25-AUG-1999; 99US-0380138.
XX PR 25-AUG-1999; 99US-0380139.
XX PR 25-AUG-1999; 99US-0380142.
XX PR 15-SEP-1999; 99US-0397342.
XX PR 18-OCT-1999; 99US-0403297.
XX PR 12-NOV-1999; 99US-0423844.
XX PR 22-AUG-2000; 2000US-0644848.
XX PR 18-SEP-2000; 2000US-0664610.
XX PR 18-SEP-2000; 2000US-0665350.
XX PR 08-NOV-2000; 2000US-0709238.
XX PR 20-DEC-2000; 2000US-0747259.
XX PR 22-MAR-2001; 2001US-0816744.
XX PR 10-MAY-2001; 2001US-0854208.
XX PR 10-MAY-2001; 2001US-0854280.
XX PR 30-MAY-2001; 2001US-0870574.
XX PR 05-JUN-2001; 2001US-0874503.
XX PR 29-JUN-2001; 2001US-0869599.
XX PR 18-JUL-2001; 2001US-0908827.
XX PR 06-DEC-2001; 2001US-0006867.
XX PA (GETH) GENENTECH INC.
XX PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski PJ;
XX PI Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
XX DR WPI; 2003-330485/31.
XX DR N-PSDB; ACA58859.
XX XX
XX FT New isolated antibody specifically binding a PRO polypeptide, useful
XX PT for the preparation of a medicament for treating disorders with the
XX PT aberrant expression or activity of the PRO polypeptide, such as tumor
XX PT conditions and cancer -
XX XX
XX PS Disclosure; Page 169-171; 406pp; English.
XX XX
XX CC The invention relates to an antibody that binds to a polypeptide with a

CC fully defined sequence given in the specification. The methods and
CC compositions (containing antibodies that specifically bind a PRO
CC polypeptide) of the present invention are useful for the preparation of a
CC medicament for the treatment of disorders associated with the aberrant
CC expression or activity of the PRO polypeptide, such as tumour conditions
CC and cancer. They can also be used to generate transgenic or knockout
CC animals useful in the development and screening of therapeutically useful
CC reagents. The PRO polypeptides and encoding nucleic acids can be used as
CC molecular weight markers for protein electrophoresis, chromosome
CC identification and tissue typing. The PRO polypeptides are useful to
CC induce angiogenesis e.g wound healing; in the treatment of sports-related
CC joint problems, articular cartilage defects, osteoarthritis or rheumatoid
CC arthritis; diabetes; hyperinsulinaemia and hypoinsulinaemia. The
CC antibodies may be used in various diagnostic, competitive binding and/or
CC immunoprecipitation assays. The present sequence represents the amino
CC acid sequence of a PRO polypeptide of the invention.
XX
SQ Sequence 596 AA;

Query Match 5.2%; Score 148.5; DB 24; Length 596;
Best Local Similarity 17.7%; Pred. No. 0.0081;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;

QY 4 GNLGNPNVNSIPPAPLPSTQDAGGRGQLNSTGPGLSRALFTPVNRSMADSGNRA 63
DB 13 GLLHLEAATNS-----NETSTANTGSSVSSG-----ASTATNSG 49

QY 64 SDVPGLPVPMRLAASEITLNDGFVLHDHGLDNLNRQIGSSV-PRVETOEDKHTAVG 122
DB 50 SSVTSSGVSTATISGSSVTSN-GVSIV-TNSEPHTTSSGISTATNSSEFTASSGISATN 107

QY 123 QRNGVETS--VLSDOEYARLQIDPEGKDKFVTGGRGAGHAMVTVASDITEARQIL 180
DB 108 SESSTSSGASTATNSE---SSTPSSGASTVNSGSSVTSSTSGASTATNSSESTVSSRAS 163

QY 181 ELLEPKGTGESKGAGSKGVGELRENSGAENTTETQTSTSTSLRSDPKLMLALGTVA 240
DB 164 TATNSSESTLSSGASTATN-SDSSTTSSGASTATNSSESTSSG-----ASTAT 211

QY 241 GLIGLAATGIVQALALTPEPDSPTTTPDAAAATATATARDQITKEAFQNPDKVNI 300
DB 212 N-----SESSTVSRASATATNSSESTT---SSGASTATNSSESTTNGAGTATNSSESTTS 264

QY 301 LG-----NAIPSGVLKDDVVANIEQAKAAGEAKQQAENNAQAOKKYDEQOAKR 351
DB 265 SGASTATNSDSTVSSGA---STATNSESTSSGAST-----ATN 302

QY 352 QEELKVSSGAGYGL-SCALILGGGIGVAVTAALHRKNQPVQTTTTTTTTTSARTVN 410
DB 303 SESSTTSSGASTATNSDSTSSGAGTATNSSESTVSSGISTVNTNSSESTPSSGATATN 362

QY 411 KPANTPAQGNVDTPGSEDTMESRRSMASTSTFFDTSSIGGPGCRIMLMKXRCMIR 470
DB 363 SESSTTSSGANTATNSSESTVSSGASTATNSSE---TTSSG----- 401

QY 471 CRLILIRLFRWIGIQISVYVSTIOHPPTDONGARLLGNPSAGIQSTVARLA----- 524
DB 402 -----VSTATNSESTTSSGASTATNSDSTSSSESTTSSSEASTATNSSESS 441

QY 525 -LSGGL-----RHDMGGLTGSNSAVNTSNPPAPGS 555
DB 442 TVSSGISTVNTNSSESTTSSGANTATNSGSSVTSAGS 477

RESULT 15
ABU72001
ID ABU72001 standard; Protein; 596 AA.
XX AC
XX ABU72001;
XX DT 11-JUN-2003 (first entry)
XX DE Novel human secreted and transmembrane protein PRO1342.

XX Human; secreted and transmembrane polypeptide;
KW chromosome mapping; gene mapping; transgenic animal; knockout animal;
KW therapeutic agent screening; chromosome identification; tissue typing;
KW gene therapy.
XX OS Homo sapiens.
XX PN US2003018183-A1.
XX PD 23-JAN-2003.
XX PF 01-MAY-2002; 2002US-0063512.
XX PR 06-DEC-2001; 2001US-0006867.
XX PA (GETH) GENENTECH INC.
XX PI Eaton DL, Filvaroff E, Gerritsen ME, Goddard A, Godowski FJ;
PI Grimaldi JC, Gurney AL, Watanabe CK, Wood WI;
XX WPI; 2003-330984/31.
DR N-ESDB; ACA60412.
XX New secreted and transmembrane PRO polypeptides and nucleic acid
XX molecules encoding the polypeptides, useful in gene therapy or
XX preparing a medicament for treating a condition that is responsive to
XX the PRO polypeptide or antibody -
XX Disclosure; Fig 100; 409pp; English.
XX The invention describes novel isolated PRO polypeptides. The PRO
XX polypeptides or anti-PRO antibodies are useful in preparing a medicament
XX for treating a condition that is responsive to the PRO polypeptide or
XX antibody. The PRO nucleotide sequences may be used as hybridisation
XX probes in chromosome and gene mapping, or in generating antisense RNA
XX and DNA. PRO nucleic acids are also useful in preparing PRO polypeptides,
XX in assays to identify other proteins or molecules involved in binding
XX reaction, to generate transgenic animals or knockout animals, which in
XX turn are useful in the development and screening of therapeutically
XX useful reagents, for chromosome identification, and tissue typing. The
XX PRO polypeptides and nucleic acid molecules are also useful in gene
XX therapy, and as molecular weight markers for protein electrophoresis
XX purposes. The anti-PRO antibodies may be used in diagnostic assays for
XX PRO, or for the affinity purification of PRO from recombinant cell
XX culture or natural sources. This is the amino acid sequence of a novel
XX human secreted and transmembrane PRO polypeptide.
XX Sequence 596 AA;
Query Match 5.2%; Score 148.5; DB 24; Length 596;
Best Local Similarity 17.7%; Pred No. 0.0081;
Matches 102; Conservative 84; Mismatches 255; Indels 135; Gaps 19;
QY 4 GNIGHNPVNNISIPAPPPLPSQTDGAGGQGLINSTGPLGSRALFPPVRNSMADSGDNRA 63
Db 13 GLLHLEAATNS-----NETSTANTGSSVSSG-----ASTATNSG 49
QY 64 SDVPGLPVNPMLAASEITLNDGFVLHDHGLDLNRLQIGSSV-FRVETQEDGKHIAVG 122
Db 50 SSVTSGVSTATIGSSVTSN-GVSIV-TNSEFHTSSGISTATNSEFSTASSGISIATN 107
QY 123 QRNGVETS--VVLSDQEVARLQSIDPEGKDKFVFTGGRGGAGHAMVTVASDITEARQRI 180
Db 108 SESSSTSSGASTATNSE-----SSTPSSGASTVNSGSSVTSNGASTATNSESTVSSRAS 163
QY 181 ELLEPKGTGSGKAGESKGVGELRNSGAENTTQTSTSTSLRSDPKLWALGTVAT 240
Db 164 TATNSESSTLSSGASTATN-SDSSTTSSGASTATNSESTSSG-----ASTAT 211
QY 241 GLIGLAATGIQVALATPEPDSFTTDDPAASATATATRDQLTKEAFQNPNDQKNIDE 300
Db 212 N-----SESTVSSRASTATNSESTT---SSGASTATNSESTTNGAGTATNSESTTS 264

QY 301 LG-----NAIPSGVLKDDVANIIEQAKAAGEAKQAIAENNAQAKKYDEQQAQR 351
Db 265 SGASTATNDSSTVSSGA---STATNSESTTSSGAST-----ATN 302
QY 352 QBELKVSSGAGYGL-SGALLILGGGIGVAVTAALHRKNQPVQETTTTTTTTTTTSARTVEN 410
Db 303 SESSTTSSGASTATNDSSTTSSGAGTATNSESSTVSSGISTVTSSESSTPSSGANTATN 362
QY 411 KPANNTPAQNVDTPGSEDTMESRRSSMASTSTFTDTSIIGGPCRCIRMLMLKHCWIRR 470
Db 363 SESSSTSSGANTATNSESTVSSGASTATNSESS---TTSSG----- 401
QY 471 CRLLILRLFRWIGIQISVYVSTIOHPPRDTTONGARLLGNPSAGIQSTYARLA----- 524
Db 402 -----VSTATNSESTTSSGASTATNDSSTTSSSEASTATNSESS 441
QY 525 -LSGGI-----RHDMGGLTGGSNSAVNTSNPPAPGS 555
Db 442 TVSSGISTVTSNESSTTSSGANTATNNGSSVTSAGS 477

Search completed: December 16, 2003, 08:52:56
Job time : 56.9828 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 48.5578 Seconds
(without alignments)
2917.574 Million cell updates/sec

Title: US-09-189-415a-2
Perfect score: 2800
Sequence: 1 MP1GNLGNVNGNHLPPAP.....GETAVSSVNAAPTGPVRFV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL 23.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rvirus.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2795	99.8	549	2	O50190
2	2676.5	95.6	550	2	O52147
3	2496.5	89.2	552	2	O9KWH9
4	2206	78.8	551	2	O68258
5	2077	74.2	547	2	O9WXK1
6	2076	74.1	547	2	O9ET11
7	1781.5	63.6	538	2	Q47016
8	1781.5	63.6	538	2	Q47014
9	1763.5	63.0	538	2	O85508
10	1528.5	54.6	558	16	Q9R396
11	1516.5	54.2	558	2	O85506
12	168.5	6.0	1158	5	O97169
13	160.5	5.7	818	6	Q9N190
14	159.5	5.7	1265	2	O9FD80
15	156.5	5.6	1323	5	O9VI63
16	156.5	5.6	1376	5	Q9BM29

17	155.5	5.6	1321	5	Q962D1	Q962d1 drosophila
18	153.5	5.5	1015	5	Q8S2W9	Q8szw9 drosophila
19	152	5.4	1203	5	Q9NSK0	Q9nsk0 caenorhabdi
20	151.5	5.4	2793	16	O8X202	O8x2q2 escherichia
21	151.5	5.4	2806	2	Q9KXA6	Q9kxa6 escherichia
22	151.5	5.4	2806	9	Q9T1K9	Q9t1k9 bacterioph
23	151	5.4	934	5	Q9VNS9	Q9vns9 drosophila
24	150.5	5.4	764	5	Q9BIT1	Q9bit1 plectreury
25	149.5	5.3	2806	9	Q9XJM1	Q9xjm1 bacterioph
26	149.5	5.3	2806	16	O8X470	O8x470 escherichia
27	149	5.3	832	2	O54356	O54356 moraxella c
28	149	5.3	2232	5	O8IFX6	O8ifx6 caenorhabdi
29	146	5.2	705	5	Q9GB22	Q9gb22 leishmania
30	146	5.2	1114	16	O8PGR8	O8pgr8 xanthomonas
31	146	5.2	1763	5	O8I110	O8i110 caenorhabdi
32	146	5.2	1829	5	O22248	O22248 caenorhabdi
33	145.5	5.2	1331	5	O8MMB1	O8mbm1 drosophila
34	145.5	5.2	3080	5	Q9V602	Q9v602 drosophila
35	145.5	5.2	3109	5	Q9BMQ0	Q9bmq0 drosophila
36	144.5	5.2	1296	5	O96506	O96506 drosophila
37	144.5	5.2	1383	5	O9VCP6	O9vcp6 drosophila
38	144	5.1	1589	6	O62672	O62672 bos taurus
39	143.5	5.1	673	11	O8K1Q4	O8k1q4 rattus norv
40	143.5	5.1	1283	5	O8TSH0	O8tsh0 anopheles g
41	143.5	5.1	1778	5	Q9NE65	Q9ne65 leishmania
42	143.5	5.1	2551	16	O8CYI8	O8cyi8 streptococc
43	142	5.1	3443	11	O8JZM8	O8jzm8 mus musculu
44	141.5	5.1	647	3	O14273	O14273 schizosacch
45	141.5	5.1	1488	16	Q8NPK0	Q8npk0 corynebacte

ALIGNMENTS

RESULT 1

ID	O50190	PRELIMINARY;	PRT;	549 AA.
AC	O50190;			
DT	01-JUN-1998 (TrEMBLrel. 06, Created)			
DT	01-JUN-1998 (TrEMBLrel. 06, Last sequence update)			
DT	01-JUN-2002 (TrEMBLrel. 21, Last annotation update)			
DE	Translocated intimin receptor.			
GN	TIR.			
OS	Escherichia coli.			
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;			
OC	Enterobacteriaceae; Escherichia.			
OX	NCBI_TaxID=562;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=E2348/69;			
RX	MEDLINE=98050926; PubMed=9390560;			
RA	Kenny B., Deviney R., Stein M., Reinscheid D.J., Frey E.A.,			
RA	Finlay B.B.;			
RT	"Enteropathogenic E. coli (EPEC) transfers its receptor for intimate			
RT	adherence into mammalian cells.";			
RL	Cell 91:511-520(1997).			
DR	EMBL; AF013122; AAB88410.1; -;			
DR	InterPro; IPR003536; Tir_receptor.			
DR	Pfam; PF03549; Tir_receptor; 1.			
DR	PRINTS; PR01370; TRNSINTMINR.			
KW	Receptor.			
SQ	SEQUENCE 549 AA; 56843 MW; 40CB8B234409A08 CRC64;			

Query Match	99.8%;	Score 2795;	DB 2;	Length 549;
Best Local Similarity	99.6%;	Pred. No. 1.1e-148;		
Matches 547;	Conservative 1;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1	MP1GNLGNVNGNHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFPLRNWADS	60	
Db	1	MP1GNLGNVNGNHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFPLRNWADS	60	
Qy	61	VDSRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDLINTQIGPSAFRVEVQADGTH	120	

```

Db 61 VDSRDI PGLPTNPRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEQADGTH 120
Qy 121 AAIGKNGLEVS VTLSPQWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEARTR 180
Db 121 AAIGKNGLEVS VTLSPQWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEARTK 180
Qy 181 ILAKLPDNDHGRQPKVDVTRSVGVSGASGIDGVDVSETHSTSTNSVRSRSDPKFVSVGA 240
Db 181 ILAKLPDNDHGRQPKVDVTRSVGVSGASGIDGVDVSETHSTSTNSVRSRSDPKFVSVGA 240
Qy 241 IAAGLAGLAATGIAQALALTPEPDPTTDPQAAANAESATKDQLTOEAFKPNQKVN 300
Db 241 IAAGLAGLAATGIAQALALTPEPDPTTDPQAAANAESATKDQLTOEAFKPNQKVN 300
Qy 301 IDANGNATPSGELXDDIVEQIAQAQKEAGEVARQAQVESNAQAQRYEDQHARRQBELQL 360
Db 301 IDANGNATPSGELXDDIVEQIAQAQKEAGEVARQAQVESNAQAQRYEDQHARRQBELQL 360
Qy 361 SSGIGYGLSSALI VAGGIGAGVTTALHRRNOPAEQTTTTTHTVVQQTGGIPQHKVLM 420
Db 361 SSGIGYGLSSALI VAGGIGAGVTTALHRRNOPAEQTTTTTHTVVQQTGGIPQHKVLM 420
Qy 421 PQERRRPSD---RRDSQGSVASTHWSDSSEVNPVYAEVGGARNSLSAHPQEEHIYDEV 480
Db 421 PQERRRPSD---RRDSQGSVASTHWSDSSEVNPVYAEVGGARNSLSAHPQEEHIYDEV 480
Qy 481 GYSVTQNFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMGLTSGGETAVSSVNA 540
Db 481 GYSVTQNFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMGLTSGGETAVSSVNA 540
Qy 541 PTPGPRFV 549
Db 541 PTPGPRFV 549

```

RESULT 2

```

ID 052147 PRELIMINARY; PRT; 550 AA.
AC 052147;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tir.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98254123; PubMed=9593291;
RA Elliott S.J., Wainwright L.A., McDaniel T.K., Jarvis K.G., Deng Y.K.,
RA Lai L.C., McNamara B.P., Donenberg M.S., Kaper J.B.;
RT "The complete sequence of the locus of enterocyte effacement (LEE)
RT from enteropathogenic Escherichia coli E2348/69.";
RL Mol. Microbiol. 28:1-4(1998).
DR EMBL; AF022236; AAC39390.1; -.
DR InterPro; IPR003536; Tir receptor.
DR Pfam; PF03549; Tir receptor; 1.
DR PRINTS; PR01370; TRNSINTMINR.
SQ SEQUENCE 550 AA; 56509 MW; 19DD08A9BE9251CB CRC64;

```

```

Query Match 95.6%; Score 2676.5; DB 2; Length 550;
Best Local Similarity 96.0%; Pred. No. 4.9e-142;
Matches 531; Conservative 4; Mismatches 11; Indels 7; Gaps 2;
Qy 1 MPIGNLGNVGNHLLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPRLNSMADS 60
Db 1 MPIGNLGNVGNHLLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPRLNSMADS 60
Qy 61 VDSRDI PGLPTNPRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEQADGTH 120
Db 61 VDSRDI PGLPTNPRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEQADGTH 120

```

```

Qy 121 AAIGKNGLEVS VTLSPQWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEARTR 180
Db 121 AAIGKNGLEVS VTLSPQWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEARTK 180
Qy 181 ILAKLPDNDHGRQPKVDVTRSVGVSGASGIDGVDVSETHSTSTNSVRSRSDPKFVSVGA 240
Db 181 ILAKLPDNDHGRQPKVDVTRSVGVSGASGIDGVDVSETHSTSTNSVRSRSDPKFVSVGA 240
Qy 241 IAAGLAGLAATGIAQALALTPEPDPTTDPQAAANAESATKDQLTOEAFKPNQKVN 300
Db 241 IAAGLAGLAATGIAQALALTPEPDPTTDPQAAANAESATKDQLTOEAFKPNQKVN 300
Qy 301 IDANGNATPSGELXDDIVEQIAQAQKEAGEVARQAQVESNAQAQRYEDQHARRQBELQL 360
Db 301 IDANGNATPSGELXDDIVEQIAQAQKEAGEVARQAQVESNAQAQRYEDQHARRQBELQL 360
Qy 361 SSGIGYGLSSALI VAGGIGAGVTTALHRRNOPAEQTTTTTHTVVQQTGGIPQHKVLM 420
Db 361 SSGIGYGLSSALI VAGGIGAGVTTALHRRNOPAEQTTTTTHTVVQQTGGIPQHKVLM 420
Qy 421 PQERRRPSD---RRDSQGSVASTHWSDSSEVNPVYAEVGGARNSLSAHPQEEHIYDEV 476
Db 421 PQERRRPSD---RRDSQGSVASTHWSDSSEVNPVYAEVGGARNSLSAHPQEEHIYDEV 476
Qy 476 AADPGYSVIVNFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMGLTSGGETAVSS 536
Db 476 AADPGYSVIVNFSGSPVTRGLIGTPGQGIQSTYALLANSGLRLGMGLTSGGETAVSS 536
Qy 537 VNAAPTGPVRFV 549
Db 537 VNAAPTGPVRFV 549

```

RESULT 3

```

ID 09KWH9 PRELIMINARY; PRT; 552 AA.
AC 09KWH9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Tir.
GN TIR.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=HK01;
RA Abe A., Nagano H.;
RT "Analyses of type III secreted proteins and Tir in enteropathogenic
RT Escherichia coli O157:H45.";
RL Submitted (DSC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB036053; BAA96815.1; -.
DR InterPro; IPR003536; Tir receptor.
DR Pfam; PF03549; Tir receptor; 1.
DR PRINTS; PR01370; TRNSINTMINR.
SQ SEQUENCE 552 AA; 57005 MW; A8D79EE22EE50A4B CRC64;

```

```

Query Match 89.2%; Score 2496.5; DB 2; Length 552;
Best Local Similarity 90.3%; Pred. No. 6e-132;
Matches 501; Conservative 11; Mismatches 34; Indels 9; Gaps 3;

```

```

Qy 1 MPIGNLGNVGNHLLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPRLNSMADS 60
Db 1 MPIGNLGNVGNHLLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPRLNSMADS 60
Qy 61 VDSRDI PGLPTNPRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEQADGTH 120
Db 61 VDSRDI PGLPTNPRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPSAFRVEQADGTH 120
Qy 121 AAIGKNGLEVS VTLSPQWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEARTR 180

```



```

Db 121 AAIKGNGLVSVALSQELQSLQSDIEGKNRFVFTGGRGGGHPMTVVASDIAEARMK 180
Qy 181 ILAKLDPDNHHGGKQPKDVTSTRSVGVSAGSIDGCVSETHSTNTSSVRSDPKFWSVGA 240
Db 181 ILAKLDPDNHHGGKQPKDVTSTRSVGVSAGSIDGCVSETHSTNTSSVRSDPKFWSVGA 240
Qy 241 IAAGLAGLAATGIAQALALATPEPDDPTTDPDQAAANAESATKDQLTQEAQORVDOHARRQELQL 360
Db 241 IAAGLAGLAATGIAQALALATPEPDDPTTDPDQAAANAESATKDQLTQEAQORVDOHARRQELQL 360
Qy 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQVESNAQAQORYEDQHARRQELQL 420
Db 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQVESNAQAQORYEDQHARRQELQL 420
Qy 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNOPAEQTTTTTHTVVOOQTGCIPOHKVAM 476
Db 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNOPAEQTTTTTHTVVOOQTGCIPOHKVAM 476
Qy 421 POERRRFS-RRDSQGSVASTHWSDSSEVVNRYAEVGGARNSLSAHQPEEHYDEV 534
Db 421 POERRRFS-RRDSQGSVASTHWSDSSEVVNRYAEVGGARNSLSAHQPEEHYDEV 534
Qy 476 AADPGYSVIIONFSGSGPVTRGLTQCGIQSTYVALLANGGLRGMGLTSGSETAV-- 537
Db 476 AADPGYSVIIONFSGSGPVTRGLTQCGIQSTYVALLANGGLRGMGLTSGSETAV-- 537
Qy 535 SSVNAAPTGPGRV 549
Db 535 SSVNAAPTGPGRV 549
Qy 538 ANANAAPTGPGRV 552
Db 538 ANANAAPTGPGRV 552

RESULT 4
O68258
ID O68258 PRELIMINARY; PRT; 551 AA.
AC O68258;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Translocated intimin receptor.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98187918; PubMed=9529069;
RA Voss E., Paton A.W., Manning P.A., Paton J.C.;
RT "Molecular analysis of Shiga toxinogenic Escherichia coli O111:H-
RT proteins which react with sera from patients with hemolytic-uremic
RT syndrome.";
RL Infect. Immun. 66:1467-1472(1998)..
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99003184; PubMed=9784578;
RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT "Translocated intimin receptors (Tir) of shiga-toxinogenic Escherichia
RT coli isolates belonging to serogroups O26, O111, and O157 react with
RT sera from patients with hemolytic-uremic syndrome and exhibit marked
RT sequence heterogeneity.";
RL Infect. Immun. 66:5580-5586(1998).
DR EMBL; AF025311; AAC69249.1; -.
DR InterPro; IPR003536; Tir receptor.
DR Pfam; PF03549; Tir receptor; 1.
DR PRINTS; PR01370; TRNSINTIMNR.
KW Receptor.
SQ SEQUENCE 551 AA; 56975 MW; 6EC95F76EF0F44CC CRC64;

Query Match 78.8%; Score 2206; DB 2; Length 551;
Best Local Similarity 79.8%; Pred. No. 1.1e-115;
Matches 442; Conservative 39; Mismatches 65; Indels 8; Gaps 3;

Qy 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGASRGAGQLINSTGALGSRLLFSPLENSTADS 60

```

```

Db 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGASRGAGQLINSTGALGSRLLFSPLENSTADS 60
Qy 61 VDSRDIPGLPTNPSRLAAATSETCLGGPEVLHDKGPLDILNTQIGPSAFRVSVDAGDTH 120
Db 61 VDSRDIPGLPVHPSRLATATSEICLLGGPEVLHDKGPLDILNTQIGPSAFRVSVDAGDTH 120
Qy 121 AAIKGNGLVSVALSQELQSLQSDIEGKNRFVFTGGRGGGHPMTVVASDIAEARMK 180
Db 121 AAIKGNGLVSVALSQELQSLQSDIEGKNRFVFTGGRGGGHPMTVVASDIAEARMK 180
Qy 181 ILAKLDPDNHHGGKQPKDVTSTRSVGVSAGSIDGCVSETHSTNTSSVRSDPKFWSVGA 240
Db 181 ILAKLDPDNHHGGKQPKDVTSTRSVGVSAGSIDGCVSETHSTNTSSVRSDPKFWSVGA 240
Qy 241 IAAGLAGLAATGIAQALALATPEPDDPTTDPDQAAANAESATKDQLTQEAQORVDOHARRQELQL 360
Db 241 IAAGLAGLAATGIAQALALATPEPDDPTTDPDQAAANAESATKDQLTQEAQORVDOHARRQELQL 360
Qy 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQVESNAQAQORYEDQHARRQELQL 420
Db 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQVESNAQAQORYEDQHARRQELQL 420
Qy 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNOPAEQTTTTTHTVVOOQTGCIPOHKVAM 476
Db 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNOPAEQTTTTTHTVVOOQTGCIPOHKVAM 476
Qy 421 POERRRFS-RRDSQGSVASTHWSDSSEVVNRYAEVGGARNSLSAHQPEEHYDEV 534
Db 421 POERRRFS-RRDSQGSVASTHWSDSSEVVNRYAEVGGARNSLSAHQPEEHYDEV 534
Qy 476 AADPGYSVIIONFSGSGPVTRGLTQCGIQSTYVALLANGGLRGMGLTSGSETAV-- 537
Db 476 AADPGYSVIIONFSGSGPVTRGLTQCGIQSTYVALLANGGLRGMGLTSGSETAV-- 537
Qy 535 SSVNAAPTGPGRV 549
Db 535 SSVNAAPTGPGRV 549
Qy 538 ANANAAPTGPGRV 552
Db 538 ANANAAPTGPGRV 552

RESULT 5
Q9WXX1
ID Q9WXX1 PRELIMINARY; PRT; 547 AA.
AC Q9WXX1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Translocated intimin receptor (Translocated intimin receptor
DE Tir).
GN Tir.
OS Escherichia coli, and
OS Citrobacter rodentium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562, 67825;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=E.coli; STRAIN=MPEC;
RA Okutani A., Itoh K., Sasakawa C.;
RT "Translocated intimin receptor (Tir) of murine pathogenic Escherichia
RT coli O15a,C:K(B).";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=C.rodentium; STRAIN=DBS100;
RX MEDLINE=21437640; PubMed=11553577;
RA Deng W., Li Y., Vallance B.A., Finlay B.B.;
RT "Locus of Enterocyte Effacement from Citrobacter rodentium: Sequence
RT Analysis and Evidence for Horizontal Transfer among Attaching and
RT Effacing Pathogens.";
RL Infect. Immun. 69:6323-6335(2001).
DR EMBL; AB026719; BAA77400.1; -.
DR EMBL; AF311901; AAL06376.1; -.

```

```

DR InterPro; IPR003536; TIR receptor.
DR EMBL; AF301617; AAG25642.1; -.
DR InterPro; IPR003536; TIR receptor.
DR Pfam; PF03549; TIR receptor; 1.
DR PRINTS; PR01370; TIRNTIMINR.
KW Receptor.
SQ SEQUENCE 547 AA; 56283 MW; 02CAC6D625FA6E1 CRC64;

Query Match 74.2%; Score 2077; DB 2; Length 547;
Best Local Similarity 76.0%; Pred. No. 1.8e-108;
Matches 418; Conservative 44; Mismatches 84; Indels 4; Gaps 3;

Qy 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNNISNNLIPAPPPLPSQTDGATRGNGSSLSISSTGLSRLLSPLRSSIVDT 60
Qy 61 VDSRDIPLCPNPSRLAAATSETCLGGFEVLHDKGPDILNTOIGPSARVEVQADGTH 120
Db 61 VDSRDVPLGPHPLRF--ATSETCLHGGFEVLHDKGPDILNTOIGPSARVEVQADGTH 118
Qy 121 AAIGKNGLEVSVTLSPOQWSSLSQIDTEGKNRFVFTGGRGGSGHPMTTVASDIAEART 180
Db 119 AAIGVKDGEVSVTLNSELQSLDTEGNGRFVFTGGRGGSGHAMVTVASDISQAREK 178
Qy 181 ILAKLPDNGHGRQPKVDTRSVGVSGASGIDGVSSETHSTTNSVRSRDPKFWVSGA 240
Db 179 IIAKLPDNGHGRQPKVDTRSVGVSGASGMDGVSETHSTTSSVRSRDPKFWVSGA 238
Qy 241 IAAGLAGLAATGIAQALALTPEPDPTTDPDQANAASATKDQTOEAFKNPENOKVN 300
Db 239 IAAGLAGLAATGIVQVALTPAPDDPTTDPDPAANAATAATKDQTKFAFQNPDKQVN 298
Qy 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQVESNAQAQRYEDQHARQELQL 360
Db 299 IDELGNAIPSGELKDDVVAQIADQAKVAGEQARQAQVESNAQAQRRHDDQAKRQELDL 358
Qy 361 SSGIGYGLSSALIIVAGGIGAGVTALHRRNPAPQTTTTTTHVVOOQTGSIPOHKVLM 420
Db 359 SSGIGYGLSSALIVGGGIGAGVTAMLHRRNPPEQ-TIATTHSVIOOQTGNTRAQGGAD 417
Qy 421 PQERRRF--DRRDSQGSVASTHSDSSSEVNPVPAEYVGGARNLSAHOPESHIYDEVAAD 479
Db 418 TTGVENASLTRDSQASVASTQWSDTSGDVNPNPAEGHMSRNNPSSLAPPEIYDEVAPD 477
Qy 480 PGYSVIQNFSGSPVTRGLITPGQGIQSTYALLANSGLRLGMLGGLTSGCTAVSVNA 539
Db 478 PNYSVIOHFGSNPNVTGLVSGPGQGIQSTYALLANSGLRLGMLGGLTGGGESAGSANA 537
Qy 540 APTGPGVRFV 549
Db 538 ATTPGVERFV 547

RESULT 6
Q9ETI1 PRELIMINARY; PRT; 547 AA.
AC Q9ETI1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Translocated intimin receptor TIR.
OS Citrobacter rodentium.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Citrobacter.
NCBI_TaxID=67825;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1843-73T, and DBS100;
RX MEDLINE=20553330; PubMed=1101562;
RA Luperchio S.A., Newman J.V., Dangler C.A., Schrenzel M.D.,
RA Brenner D.J., Steigerwalt A.G., Schauer D.B.;
RT "Citrobacter rodentium, the Causative Agent of Transmissible Murine
RT Colonic Hyperplasia, Exhibits Clonality: Synonymy of C. rodentium and
RT Mouse-Pathogenic Escherichia coli.";
RL J. Clin. Microbiol. 38:4343-4350(2000).

DR EMBL; AF301618; AAG40758.1; -.
DR EMBL; AF301617; AAG25642.1; -.
DR InterPro; IPR003536; TIR receptor.
DR Pfam; PF03549; TIR receptor; 1.
DR PRINTS; PR01370; TIRNTIMINR.
KW Receptor.
SQ SEQUENCE 547 AA; 56270 MW; CB8318B301049C37 CRC64;

Query Match 74.1%; Score 2076; DB 2; Length 547;
Best Local Similarity 76.0%; Pred. No. 2.1e-108;
Matches 418; Conservative 44; Mismatches 84; Indels 4; Gaps 3;

Qy 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNNISNNLIPAPPPLPSQTDGATRGNGSSLSISSTGLSRLLSPLRSSIVDT 60
Qy 61 VDSRDIPLCPNPSRLAAATSETCLGGFEVLHDKGPDILNTOIGPSARVEVQADGTH 120
Db 61 VDSRDVPLGPHPLRF--ATSETCLHGGFEVLHDKGPDILNTOIGPSARVEVQADGTH 118
Qy 121 AAIGKNGLEVSVTLSPOQWSSLSQIDTEGKNRFVFTGGRGGSGHPMTTVASDIAEART 180
Db 119 AAIGVKDGEVSVTLNSELQSLDTEGNGRFVFTGGRGGSGHAMVTVASDISQAREK 178
Qy 181 ILAKLPDNGHGRQPKVDTRSVGVSGASGIDGVSSETHSTTNSVRSRDPKFWVSGA 240
Db 179 IIAKLPDNGHGRQPKVDTRSVGVSGASGMDGVSETHSTTSSVRSRDPKFWVSGA 238
Qy 241 IAAGLAGLAATGIAQALALTPEPDPTTDPDQANAASATKDQTOEAFKNPENOKVN 300
Db 239 IAAGLAGLAATGIVQVALTPAPDDPTTDPDPAANAATAATKDQTKFAFQNPDKQVN 298
Qy 301 IDANGNAIPSGELXDDIVEQIAQAQKEAGEVARQQAQVESNAQAQRYEDQHARQELQL 360
Db 299 IDELGNAIPSGELKDDVVAQIADQAKVAGEQARQAQVESNAQAQRRHDDQAKRQELDL 358
Qy 361 SSGIGYGLSSALIIVAGGIGAGVTALHRRNPAPQTTTTTTHVVOOQTGSIPOHKVLM 420
Db 359 SSGIGYGLSSALIVGGGIGAGVTAMLHRRNPPEQ-TIATTHSVIOOQTGNTRAQGGAD 417
Qy 421 PQERRRF--DRRDSQGSVASTHSDSSSEVNPVPAEYVGGARNLSAHOPESHIYDEVAAD 479
Db 418 TTGVENASLTRDSQASVASTQWSDTSGDVNPNPAEGHMSRNNPSSLAPPEIYDEVAPD 477
Qy 480 PGYSVIQNFSGSPVTRGLITPGQGIQSTYALLANSGLRLGMLGGLTSGCTAVSVNA 539
Db 478 PNYSVIOHFGSNPNVTGLVSGPGQGIQSTYALLANSGLRLGMLGGLTGGGESAGSANA 537
Qy 540 APTGPGVRFV 549
Db 538 ATTPGVERFV 547

RESULT 7
Q47016 PRELIMINARY; PRT; 538 AA.
AC Q47016;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-AUG-1999 (TrEMBLrel. 11, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Translocated intimin receptor (Tir).
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=REPEC 83/39;
RA O'Brien R.A.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

```



```

[4]
RP SEQUENCE FROM N.A.
RC STRAIN=B10;
RA Marches O., Nougayrede J.-P., Boury M., Mainil J., Charlier G.,
RC Boullier S., De Rycke J., Milon A., Oswald E.;
RA "Role of Tir and intimin in the pathogenesis of rabbit
RT enteropathogenic Escherichia coli.";
RL Submitted (DSC-1998) to the EMBL/GenBank/DBJ databases.
[5]
RP SEQUENCE FROM N.A.
RC STRAIN=413/89-1;
RA Benkel P., Chakraborty T.;
RT "Genetic organisation and sequence of the LEE II locus in Shiga toxin-
RT producing Escherichia coli.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U59502; AAC32028.2; -.
DR EMBL; AJ223063; CAAL1065.1; -.
DR EMBL; AF070068; AAC69316.1; -.
DR EMBL; AF132728; AAD27868.1; -.
DR EMBL; AF113597; AAP03080.1; -.
DR EMBL; AJ277443; CAC81869.1; -.
DR InterPro; IPR005356; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTIMNR.
DR KW Hypothetical protein; Receptor.
SQ SEQUENCE 538 AA; 55420 MW; 31D7A8E227B3D06C CRC64;

Query Match 63.6%; Score 1781.5; DB 2; Length 538;
Best Local Similarity 65.4%; Pred. No. 6.3e-92;
Matches 367; Conservative 66; Mismatches 93; Indels 35; Gaps 11;

Qy 1 MPIGNLNNVNGNHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLPRLNSMADS 60
Db 1 MPIGNLGNPNVRALIPAPPLPSQTDGAG-GARNQLNSNGPMGSRLLTFPIRNSVADA 59

Qy 61 VDSR--DIPGLTPNPSRLAAATSETCLLGGFEVLHDKGFLDILNTOIGPSAFRVEQADG 118
Db 60 ADSRASDIPGLTPNPLRFAA--SEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDDG 117

Qy 119 THAAIGKNGLEVSVTLSPQWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEAR 178
Db 118 SHVAIGKNGLETTTVLSEQFSSLSQSLDPEGKNKFVFTGGRGGAGHAMVTVASDIAEAR 177

Qy 179 TRIILAKLDP-DNHGGRQPKVDVTRSVGVGSAGIDGVSSTHTSTTSSVRSRDPKFWVS 237
Db 178 QRIIDKLEPKDTKETKEPDGPN-----GEGKIIIEHTSTSTSSLRADPKLWLS 226

Qy 238 VGAIAGLAGLAATGIAQALALTPEDDPTTTPDPAANAASATKQDLTOEAFKNPNQ 297
Db 227 LGTIAAGLIGMAATGIAQAVALTPEPDPTTTPDPAANAATAEAAKQDLTKEAFQNPQ 286

Qy 298 KVNIDANGNAIPSGELXDDIVEQIAQAKAGEVARQOAVESNAQAOQRYEDOHARRQEE 357
Db 287 KVNIDENGNAIPSGELKDDVVAQIAEQAKAGEARQAEIENSQAOKYDEQHAKEQEE 346

Qy 358 LQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHVQQQTGGIPQHKV 417
Db 347 MSLSSGVGYGLSGALILGGGIGAGVTAALHKNQPAEQ--TITRTVVNQ-----PTNNA 400

Qy 418 ALM-----POERRRFRDRRDSQGSVASTHWSRSS--EVNVPYAEVGGARNLSAHQP 468
Db 401 SAQGNVDTSGPEESP--ASRRNSNASLAS-NGSDTSSTGTVPENYADVGMPRNDSLARIS 457

Qy 469 EEHIYDEVAADPGYSVIQNFSGSPVTRGLIGTTCGGIQTSTYALLANGSGRLGCGGLTS 528
Db 458 EEPYIYDEVAADPNYSVIQHFSGNSPVTGRVLTGTPGGIQTSTYALLANGSGRLGCGGLTG 517

Qy 529 GGETAVSSVNAAPTGPVRFV 549
Db 518 GGESAVSTANAAPTGPAPRFV 538

```

```

085508 PRELIMINARY; PRT; 538 AA.
AC 085508;
DT 01-NOV-1998 (T-EMBLrel. 08, Created)
DT 01-NOV-1998 (T-EMBLrel. 08, Last sequence update)
DT 01-JUN-2002 (T-EMBLrel. 21, Last annotation update)
DE Translocated intimin receptor Tir.
GN Tir.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EPEC87A;
RX MEDLINE=99003184; PubMed=9784578;
RA Paton A.W., Manning P.A., Woodrow M.C., Paton J.C.;
RT "Translocated intimin receptors (Tir) of shiga-toxicogenic Escherichia
RT coli isolates belonging to serogroups O26, O111, and O157 react with
RT sera from patients with hemolytic-uremic syndrome and exhibit marked
RT sequence heterogeneity.";
RL Infect..Immun. 66:5580-5586(1998).
DR EMBL; AF070069; AAC69318.1; -.
DR InterPro; IPR005356; Tir_receptor.
DR Pfam; PF03549; Tir_receptor; 1.
DR PRINTS; PR01370; TRNSINTIMNR.
DR KW Receptor.
SQ SEQUENCE 538 AA; 55602 MW; 447052A0E3214D6D CRC64;

Query Match 63.0%; Score 1763.5; DB 2; Length 538;
Best Local Similarity 64.5%; Pred. No. 6.5e-93;
Matches 362; Conservative 71; Mismatches 91; Indels 35; Gaps 11;

Qy 1 MPIGNLNNVNGNHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLPRLNSMADS 60
Db 1 MPIGNLGNPNVRALIPAPPLPSQTDGAG-GARNQLNSNGPMGSRLLTFPIRNSVADA 59

Qy 61 VDSR--DIPGLTPNPSRLAAATSETCLLGGFEVLHDKGFLDILNTOIGPSAFRVEQADG 118
Db 60 ADSRASDIPGLTPNPLRFAA--SEVSLHGALEVLHDKGGLDTLNSAIGSSLFRVETRDDG 117

Qy 119 THAAIGKNGLEVSVTLSPQWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEAR 178
Db 118 SHVAIGKNGLETTTVLSEQFSSLSQSLDPEGKNKFVFTGGRGGAGHAMVTVASDIAEAR 177

Qy 179 TRIILAKLDP-DNHGGRQPKVDVTRSVGVGSAGIDGVSSTHTSTTSSVRSRDPKFWVS 237
Db 178 QRIIDKLEPKDTKETKEPDGPN-----GEGKIIIEHTSTSTSSLRADPKLWLS 226

Qy 238 VGAIAGLAGLAATGIAQALALTPEDDPTTTPDPAANAASATKQDLTOEAFKNPNQ 297
Db 227 LGTIAAGLIGMAATGIAQAVALTPEPDPTTTPDPAANAATAEAAKQDLTKEAFQNPQ 286

Qy 298 KVNIDANGNAIPSGELXDDIVEQIAQAKAGEVARQOAVESNAQAOQRYEDOHARRQEE 357
Db 287 KVNIDENGNAIPSGELKDDVVAQIAEQAKAGEARQAEIENSQAOKYDEQHAKEQEE 346

Qy 358 LQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHVQQQTGGIPQHKV 417
Db 347 MSLSSGVGYGLSGALILGGGIGAGVTAALHKNQPAEQ--TITRTVVNQ-----PTNNA 400

Qy 418 ALM-----POERRRFRDRRDSQGSVASTHWSRSS--EVNVPYAEVGGARNLSAHQP 468
Db 401 SAQGNVDTSGPEESP--ASRRNSNASLAS-NGSDTSSTGTVPENYADVGMPRNDSLARIP 457

Qy 469 EEHIYDEVAADPGYSVIQNFSGSPVTRGLIGTTCGGIQTSTYALLANGSGRLGCGGLTS 528
Db 458 EEPYIYDEVAADPNYSVIQHFSGNSPVTGRVLTGTPGGIQTSTYALLANGSGRLGCGGLTG 517

Qy 529 GGETAVSSVNAAPTGPVRFV 549
Db 518 GGESAVSTANAAPTGPAPRFV 538

```



```

QY 1 MPIGNLGNVNGNHL1PPAPPLPSQTDGAARGGTGHLISSTGALGSSLSFPLRNMS 60
DB 1 MPICNLGNHNPVNNV1PPAPPLPSQTDGA--GGRQLINSTPLGSRALTPVRNMS 58
QY 61 VDSR--DIPGLTPNPSRLAATSETCLLGGFEVLHDKGPLDILNTOIGPSAFRVEQADG 118
DB 59 GDNRASDPVGLPVNPMRLAA--SBITLNDGFEVLHDKGPLDILNTOIGSSVFRVETQEDG 116
QY 119 THAAIGKNGLEVSVTLSPQEWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEAR 178
DB 117 KHIAVGORNGVETSVLSDQYARLQSLDPKQKRFVFTGGRGAGHAMTVVASDITEAR 176
QY 179 TRILAKLDPDNNHGRQPKVDVTRSVGVGSAGI-----DDGV--VSETHSTTNNSVRS 230
DB 177 QRILELLEPKGTG-----ESKGAGESKGVGELRESNSGAENTTETOTSTSTSLRS 227
QY 231 DPKFVSVGAIAAGLAGLAATGIAQALATPEPPDDPTTTPDQAAANAASATKQDLTOEA 290
DB 228 DPKLWALGTATGLIGLAATGIVQALATPEPPDPTTTPDAAAASETATRDQLTKEA 287
QY 291 FKNPENQKVNIDANGNATPSELXDDIVEQIAQQAQAGEVARQQAQVESAQOQRYEDQ 350
DB 288 FQNDPNQKVNIDELGNATPSGVLKDDVVANIEEQAKAGEAKQQALENNAQAQKQYDEQ 347
QY 351 HARRQEELQSSGIGYGLSSALIIVAGGIGAGVTTALHRRNQAPBQTTTTT-----HT 403
DB 348 QAKRQEELKVSSGAGYGLSGALILGGGIGVAVTAALHKKQNPVQTTTTTTTTTTSART 407
QY 404 VVQQTGGIPOHKVALMPQERRRSDRRDSOGSVASTHWSDS--SEVVNPVAVGVGGHNS 462
DB 408 VENKANNTPAQGNVDTPGSEDTMESRRSSMASTLSTFFDTSSTGTQVNPYADV-----K 462
QY 463 LSAHQPEEHIYDEVAADPGYVIONFGSGGPVT-----GRLLGTPGOGIQ 507
DB 463 TSLHD-----SQVPTSNTSVQMGNTSDSVVTSIQHPRPRTDNGARLLGNPSAGIQ 516
QY 508 STYALLANGSGIRLGMWGLTSGGTAVSSVNAAPTGPVRV 549
DB 517 STYARLALRGGLRHDGMGLTGGSNAVNTSNPPAGSHRFV 558
RESULT 12
Q97169
ID Q97169 PRELIMINARY; PRT; 1158 AA.
AC Q97169;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Hairless protein.
GN H OR HAIRLESS.
OS Drosophila hydei (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7224;
RN [1]
RP SEQUENCE FROM N.A.
RA Marquart J., Alexief-Damianof C., Preiss A., Maier D.;
RT "Rapid divergence in the course of Drosophila evolution reveals
RT structural important domains of the Notch antagonist Hairless."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ232783; CAB38221.1; --
DR FlyBase; FBgn0019190; DbydVH.
SQ SEQUENCE 1158 AA; 119635 MW; 0B611D3739CF397F CRC64;
Query Match 6.0%; Score 168.5; DB 5; Length 1158;
Best Local Similarity 18.9%; Pred. No. 0.39;
Matches 130; Conservative 89; Mismatches 238; Indels 231; Gaps 26;
QY 6 LGNVNNGNHL-----IPAPPLPSQTDGAARGGTGHLISSTGALG 45
DB 57 LKNGINGTNTAAAAAATLTTAAAAAVTQLPAPELKTTSATAANATNTSTSSSTATS 116

```

```

QY 46 SRSLSFPLRNMSADSVDSRDIPGLPTNPSRLAATSETCLL--GGFEVLHDKGPLDILNTQ 104
DB 117 TSGNTTTPALSTKTNNSSNS-----SSSLIMATASATALVAGGAGVTAAPKAPADYMAVG 171
QY 105 IGPSAFRVEQADGTHAAIGKNGLEVSVTLSPQEWSSLSQSDTEGKRFVFTGGRGSG-- 163
DB 172 L-----DYSALNSPNATKSSKFFVGTGNGSSSF 198
QY 164 ---GHPMTVASD-----IAEARTRILAKLDPDNNHGG-----ROPKDV 198
DB 199 DMGRHPISMRSNNMSGYGRLQFFKDGKFILE-----LARAQDGKGGWVSVPRKPFRT 253
QY 199 DTRSGVGVGSAGIDDDGVVSETHT-----STNSSVRSDD-----KFWVSVGAIAAG 244
DB 254 PSAAT---SATVTPTSATVTTTPKNNSTLSLSDSDNSSTQSSPQWQDOPWKQT----- 304
QY 245 LAGLAATGIAQALAL-----TPEDDDPT--TTDPDQAAANAASAT 282
DB 305 ---RPRRGISKELUSLYQRPVNVLSQSAQAAAKRRRPRPYEPTIATENHQSIFERVSAQ 361
QY 283 KQDLTOEAFAFNPNQKVNIDANGNATPSELXDDIVEQIAQQAQAGEVARQQAQVESAQ 342
DB 362 EN--GDETLLKAVEVDTKAPEIQSDDIKSEIKELTEGKAKSDSKACSSDNKDLNDSAK 419
QY 343 AQORYEDQHARQEE--LQSSGIGYGLSSALIIVAGGIG-----AGVTTA-----LHR 388
DB 420 AEATLEKMNSTSEDAAMSVTDGAPTTSCLVNGNTSEVNGEPKAKTKCETASEKDTTR 479
QY 389 RNQAPBQTTTTT-----HTVVQQTGGIPOH-----KVALMPQ 422
DB 480 TGTPTATKTMTMDIRLKQKQPRAKLNSIIQKLDGVPARLEQLSKTPAVAAATASAVAS 539
QY 423 ERRFRSDRRDSOGSVASTHWSDSSEVVNPVAVGVGGHNSLSAQHPEEHI-----Y 473
DB 540 TADRISSNSSGGGAGISL--SHSLAHKVPSPSSAAAAASRLVEYHHHHQHVSPKRILREF 597
QY 474 DEVAADPGYVIONFGSGGPVTGRLIGTPQGGIQTSTYALLANGSGIRLGMWGLTSGGTAV 531
DB 598 EKVSLEDNNGCVNNGSGVG-----GGISGGAGGKRRAKGS 633
QY 532 TAVSSV-----NAAP-----TPGP 545
DB 634 TATSAVTTKSMNPINLAPPQAKVLSTPTP 661
RESULT 13
Q9NIP0
ID Q9NIP0 PRELIMINARY; PRT; 818 AA.
AC Q9NIP0;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Submaxillary mucin (Fragment).
GN BSM1.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=breed Angus;
RC MEDLINE=2023253; PubMed=10759843;
RX Jiang W., Gupta D., Gallagher D., Davis S., Bhavanandan V.P.;
RT "The central domain of bovine submaxillary mucin consists of over 50
RT tandem repeats of 329 amino acids: chromosomal localization of the
RT BSM1 gene and relations to ovine and porcine counterparts."
RL Eur. J. Biochem. 267:2208-2217(2000).
DR EMBL; AF178428; AAF67279.1; --
FT NON_TER 1
FT NON_TER 818
SQ SEQUENCE 818 AA; 73014 MW; B5BB44F84F66F86B CRC64;

```

Query Match		5.7%; Score 160.5; DB 6; Length 818;
Best Local Similarity		21.6%; Pred. No. 0.68;
Matches 122; Conservative		64; Mismatches 221; Indels 157; Gaps 21;
QY	10 VNGNHLPPAPPLPSQTDGAAGGTGHLISSTGALGSRSLFSLRNSMADSVDSRDIPLGL	69
DB	174 VSGAAPSPPGSPKRSRATVSGGSOPTVALSGATGTSAGPSTRST-----SSAIPAT	227
QY	70 PTNPSRLAAATSETCLLGGFEVLHDKGDLILNTQIGPSAFRVEVQADGTHAIGKNGKL	129
DB	228 PGSTTGRAAGATPA-----VDSQQTASLPAARLTALPGPTISAPGETSES	274
QY	130 EVSVTLSPQWSSLSQIDTEGKNRFVFTGGRGSGHGMVTVASDIABARTRI-----LAKL	185
DB	275 RSSVP-----CGSETTQPG-----AGSEPTLSPGVTRTTALRGSETRKVPSTGVSGL	322
QY	186 DPNHGRQPKVDVTVSGVSGSAGIDDDGVVSETHST-----TNSSVSRDPKFWVSVAIA	242
DB	323 PGSTQGG-----SAATGGSGAGSGPTAPVSGETRTSVISGTNPVSGAP---VTPGSSA	373
QY	243 -----AGLAGLAATGIAQALALTPEDPTTDPDQAANAASATK	283
DB	374 GSSGAPGTGPGSETASPLSAGATGTSATGTSI-----PPSGAPVTPPPLISTGA-SAGP	429
QY	284 DLTQEAFAKPNQKVNIDANGNAIP--SGELXDDIVEQIAQQAKEAGEVAROQAVESNAQ	342
DB	430 PASSESTVTLPGATGTDVLRSGTSLPVSG-----GAVTPAPSPGSSA-----	472
QY	343 AQRVEDQHARQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTH	402
DB	473 -----TAGPGVGSATTVOASGATGADVL-----	495
QY	403 TVVQOQTGGIPQHKVALMP--QERRRFSRRDSQGSVASTHSDSSSEVVNPVAVESGAR	460
DB	496 -----RSTSLPVGVAIVSPSSPGRSATVSSQSGOPTVALSGATGTSVGP-----SGTR	547
QY	461 NLSAHOPEHIYDEVAADPGYSVQNFGSGPVTGRIGTGGIQTSTYALLANSGLR	520
DB	548 FSSSA-----IPATPG-----SITGRAAG--ACTPGVDSQQTASLPAARPTA	588
QY	521 LGMGLTSGGETAVSSVNNAPTGP	544
DB	589 LGPGTSAPSGET---SESRSVPG	609
RESULT 14		
ID	Q9FDA0	PRELIMINARY; PRT; 1265 AA.
AC	Q9FDA0;	
DT	01-MAR-2001 (TrEMBLrel. 16, Created)	
DT	01-MAR-2001 (TrEMBLrel. 16, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	Putative outer membrane protein Xada.	
GN	XADA.	
OS	Xanthomonas oryzae (pv. oryzae).	
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;	
OC	Xanthomonadaceae; Xanthomonas.	
OX	NCBI_TaxID=64187;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RA	Ray S.K., Rajeshwari R., Sonti R.V.;	
RT	"A putative outer membrane protein from Xanthomonas oryzae pv. oryzae that is involved in virulence.";	
RL	Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.	
DR	EMBL; AF288222; AAC01335.1; --	
DR	InterPro; IPR005594; Yada.	
DR	Pfam; PF03895; Yada; 1.	
SQ	SEQUENCE 1265 AA; 119856 MW; 21762579B5EC70A2 CRC64;	
Query Match		5.7%; Score 159.5; DB 2; Length 1265;
Best Local Similarity		23.4%; Pred. No. 1.4;
Matches 136; Conservative		73; Mismatches 229; Indels 143; Gaps 29;
QY	3 IGLNNGNVNNGNHLPPAPPLPSQTDGAAGGTGHLISS--TGALGSRSLFSLR-----N 55	
DB	592 LGNTARALGDNSSAVGNSAVASDVGATANGAGAQAALSAVTIALGSKAVASDLQAIABGFN	651
QY	56 SMADSVDSRDIPLGPTNPSRLAAATSETCLLGGFEVLHDKGDLILNTQIGPSAFRVEVQ	115
DB	652 STASNVGSTALGDFGSSGRLSSALGYSAVASSVD-----STAVG-----VAAQ	695
QY	116 ADGTHA-AIGE--KNGLENSVTLSPQWSSLSQIDTEGKNRFVFTGGRGSGHGMVTVAS	172
DB	696 ATGVSVAIVGETSKATGEESVAVGGAFSGWIPTQASGKAAGAFAGAWATADYTTAIGR	755
QY	173 D-IAEA--RTRILAKLD--PDN-----HGGRPQKDVDTSTRVGV--GSAAGIDD-----GV	215
DB	756 DSVADGWNATAVGQSADALADNTLALGGGSRARAKAVGASVIGVDASATGINSTGVGRQVNV	815
QY	216 VSETHST--TNSSVSRDPKFWVSVAIAAGLAGLAATGIAQALALTPEDPTTDPDQA	274
DB	816 IGENAVSVGVNSYVROS-----AVNGVALG-ANAGATG-ADSVALL---GSGSSTYDADTV	865
QY	275 ANAASATKDQLTOEAFKPNQKVN-----DANGNAIPSGELXDDIVEQIAQQAKEA	328
DB	866 SVSGNGRGGPATRI-----VNGAGAVASASTDALNGQLFESL-----	906
QY	329 GEVAROQAVESNAQAQORYEDQHARQEELQLSSGIGYGLSSALIVAGGIGAGVTTALH-	387
DB	907 -----SNA-----ASFLLGG--AAIGAQGVFAVPTVLYIQG	934
QY	388 --RRNQPAEQTTTTTHTVVQOQTGGIPQHKVALMPQERRRFSRRDSQGSVASTHSDS	445
DB	935 ASYNNVGAALTALDSKVTELDARSQGTTPANTAA-----RTASLTATVPAAMAATAVS	988
QY	446 SSSV-----VNPVAVESGARNSLSAHOPEHIYDEVAADPGYSVQNFGSGSPVTGR	501
DB	989 SSVNSTATDATAGVGT-----PTAVGSIPTAAISTVGTAAVANNVTGTGTAIG-	1039
QY	502 PGQGIQSTYALLANSGL-----RLGMGLTS--GGETAVSSV	537
DB	1040 -----GSAYAHGANDTAIGSNARVNADGSTAVGANTQIAAV	1075
RESULT 15		
ID	Q9VI63	PRELIMINARY; PRT; 1323 AA.
AC	Q9VI63; Q9NHX4;	
DT	01-MAY-2000 (TrEMBLrel. 13, Created)	
DT	01-OCT-2002 (TrEMBLrel. 22, Last sequence update)	
DT	01-MAR-2003 (TrEMBLrel. 23, Last annotation update)	
DE	CG1070 protein (Alhambra).	
GN	ALHAMBRA OR CG1070.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;	
OC	Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=Berkeley;	
RA	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	
RA	Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,	
RA	Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,	
RA	Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,	
RA	Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,	
RA	Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,	
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolehakov S.,	
RA	Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,	
RA	Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,	
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,	
RA	de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,	

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista A.C., Ferraz C., Ferrera S., Fleischmann W.,
RA Fobler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.B., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M.B., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.A., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA Carlson J.W., Center A., Champs M., Davenport L.B., Dietz S.M.,
RA Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA Ferrera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA Ibegwan C., Jalali M., Kruse D., Li P., Mattei S., Moshrefi A.,
RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA Phouanavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT "Sequencing of *Drosophila melanogaster* genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
RA Tupy J.L., Bergman C., Bernier B., Carlson J.W., Celnik S.E.,
RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA Seale S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT "Annotation of *Drosophila melanogaster* genome."
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Adams M.D., Celnik S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [5]
RP SEQUENCE FROM N.A.
RA FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN [6]
RP SEQUENCE FROM N.A.
RA Perrin L., Dura J.M.;
RT "Alhambra, a *Drosophila* homolog of mammalian AP-10 and AP-17."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF003672; AAF54062.2; -
DR EMBL; AF217960; AAF72595.1; -
DR FlyBase; Fgn0037471; Alhambra.
DR InterPro; IPR00104; Antifreeze_1.
DR InterPro; IPR001965; Znf_PHD.
DR Pfam; PF00628; PHD; 1.
DR PRINTS; PR00308; ANTIFREEZE1.

DR SMART; SM00249; PHD; 2.
DR PROSITE; P550016; ZF_PHD_2; 1.
SQ SEQUENCE 1323 AA; 132884 MW; D53C0C8AF392F9A6 CRC64;
Query Match 5.6%; Score 156.5; DB 5; Length 1323;
Best Local Similarity 20.2%; Pred. No. 2.2; Indels 177; Gaps 25;
Matches 135; Conservative 85; Mismatches 270;
QY 3 IGLNNVNGNHLIPPPALPSPQDTGAARGTGHLS-----STGALGRSLFSP-LRNS 56
DB 322 IGNISSLNN---LPGSSSTSSAGNVPFGSGGIISASSGGATQTSQSSTAPGTTKS 378
QY 57 MADVSDRDIPLPTNPRLAATSTCLLGGFEVLDKGLDILNTQIGPSAFRVEVOA 116
DB 379 SASSSSSSSSYKEKHSKLSKSTSSKDKDG-----KDSSTNSANNPNTNNSASSTSSNS 433
QY 117 DGHAAIGKEKNGLEVS-----VTLS PQEWSLSQIDTE-GKRVFVTGGRGSGG-HPM 167
DB 434 SSTR-----EKSSSKLSKNKDSNOVPSATSLSTSSINTQPSSTSTATAGSGGTGTHVS 489
QY 168 VTVASDIAEARTILAKLDPDNHHGRQPKDQVDTRSVGVSGASGIDDG----- 215
DB 490 SSAASGLNSAPS-----TTNEHSNHAHTLSTNGTGAGSAACKLQSVLSNLSGSGFGS 542
QY 216 -----VSETHSTTNSV-----RSDPKFWVSVGAIAAGLAGLAA--TGIAQALATPE 262
DB 543 DLRSVSTSSSTVNDSTGGFGSNSGERENLSGAGSSASNNMPTIAPGTGCVSSAAT-- 600
QY 263 PDPTTTDPQANAAEATKDLQTOEAFKNPNQKNVID---ANGNAIPSGELXDDIVE 319
DB 601 --NLSTNKGSSSTANSLTSTSTSGSSNSSSSKRKADSAKSTSISSISGSALEDNNS 658
QY 320 QIAQ-QAKEAGEVARQOAVESNAQAQOYEDQHARQEEL-----QLSSSGIGYGLSSA 371
DB 659 LISRYDIKDV-----QVALTPLTDFEKEIEKSKRQRTLSPPHTQTSATAEVNAPLASS 713
QY 372 LIVAGGIGAGVTALHRRNOPAQTTT-----TTHVTVVQQQTG 410
DB 714 TSASIAVTASATAA---SAPPATGTTTLGSSISGNAGSTSSGNSGVSATGGAQSAVSGG 770
QY 411 GIPQHKVALMPQERRRFRDRDSQSVASTHWD-----SSSEVVNPFYAE 455
DB 771 GYPKTESKSGTASAGSGSGSSNTSTTKHGNIKDIISSSSNQQAATASSSSAPSLYS 830
QY 456 V-----GGARNSLSAHQPEEHIYDEVAADP----- 480
DB 831 VPLSTANVPGINLPTSTSTSTSESHSASSRSSGAGSQHQQLSNALVGPMPGGSAGAF 890
QY 481 -----GYSVIQNFSG-SGPVTGRLIGTPGQGIQSTYALLANSGLRLMGKG--LTS 529
DB 891 HGGTTTSAGSSSVIQQQSGKSPALGTLVS-----GNSGSGSIISASGFLPSPG 937
QY 530 GETAVSS 536
DB 938 NLTATT 944

Search completed: December 16, 2003, 08:55:42
Job time : 52.5578 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 21.306 Seconds
(without alignments)
1211.758 Million cell updates/sec

Title: US-09-189-415A-2

Perfect score: 2800

Sequence: 1 MPIGNLGNVNGNHLIPAP.....GETAVSSVNAAPTGPVRFV 549

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	157.5	5.6	494	1	SOX9_CHICK
2	150.5	5.4	1608	1	H1YA_SERWA
3	150	5.4	1005	1	Y456_CHLTR
4	144	5.1	883	1	E74B_DROME
5	140	5.0	2132	1	PGCA_MOUSE
6	137	4.9	1532	1	IGA_NEIGO
7	136	4.9	1569	1	YFJA_ECOLI
8	135.5	4.8	2124	1	PGCA_RAT
9	134.5	4.8	1150	1	APMU_PIG
10	132.5	4.7	1025	1	SLAP_CAUCK
11	132	4.7	666	1	FLID_VIBCH
12	132	4.7	827	1	CSG_HALVO
13	131.5	4.7	618	1	ZEST_DROVI
14	130.5	4.7	940	1	MAZ4_SCHCO
15	130	4.6	1260	1	ALSI_CANAL
16	130	4.6	1596	1	MAM_DROME
17	130	4.6	2703	1	ACTC_DROME
18	129	4.6	1341	1	ACIN_HUMAN
19	129	4.6	2415	1	PGCA_HUMAN
20	128	4.6	827	1	XANP_XANG2
21	127.5	4.6	509	1	SOX9_PIG
22	127	4.5	784	1	SP4_HUMAN
23	126.5	4.5	2493	1	CYAA_USTWA
24	125.5	4.5	673	1	Y552_HUMAN
25	125	4.5	2442	1	CBP_HUMAN
26	124	4.4	977	1	BAB1_DROME
27	124	4.4	1589	1	PHF_DROME
28	123.5	4.4	785	1	SP1_HUMAN
29	123.5	4.4	829	1	E74A_DROME
30	123.5	4.4	2464	1	MAPB_MOUSE
31	123.5	4.4	3591	1	FIAB_BORPE
32	123	4.4	980	1	SVN_DROME
33	122.5	4.4	1227	1	J1P3_DROME

ALIGNMENTS

RESULT 1

SOX9_CHICK
ID SOX9_CHICK STANDARD; PRT; 494 AA.
AC P48434;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Transcription factor SOX-9.
GN SOX9.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RA Uvanogho D., Rex M., Cartwright B.J., Pearl G., Scotting P.J.,
RA Sharpe P.T.;
RC Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RL -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: Contains 1 HMG box domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; U12533; AAB09663.1; -.
DR HSSP; P48436; 1SX9.
DR InterPro; IPR000910; HMG_12_box.
DR Pfam; PF00505; HMG_box; 1.
DR SMART; SM00338; HMG_1.
DR PROSITE; PS0116; HMG_BOX_2; 1.
KW Transcription regulation; DNA-binding; Nuclear protein.
FT DNA_BIND 106 174 HMG_BOX.
FT DOMAIN 215 220 POLY-SER.
FT DOMAIN 342 360 GLN/PRO-RICH.
FT DOMAIN 394 401 POLY-GLN.
SQ SEQUENCE 494 AA; 54942 MW; AF41D483B5016349 CRC64;

Query Match 5.6%; Score 157.5; DB 1; Length 494;

Best Local Similarity 18.9%; Pred. No. 0.031;

Matches 74; Conservative 48; Mismatches 142; Indels 127; Gaps 11;

Qy 103 TOIGPSAFRVEVDAGTHAATGEKNGLSEVSTLSPQEWSSLSQSDTEGKNRFFVTGGGG 162

Db 197 THISFNAIFKALQAD-----SPQSSSSISEVHSPGEH-----SQSQG 234

Qy 163 SCHPMVTVASDIARTILAKLPDNDHGGCHQPKDVTDSVGVGSASIDGVVSEHTS 222

Db 235 PPTPTPTTKTDAQQPGKDLKREGFLAEGGGPHIDFRVDYDIGELS---SDVISNIETFF 291

Qy 223 TTNS-----SVRSDFKFWVSVAIAAGLAGLAATGTAQALALTPSPDDPTT 268

34 122.5 4.4 1306 1 MSB2_YEAST P22334 saccharomyc
35 122.5 4.4 2541 1 TLN1_MOUSE P26039 mus musculu
36 121.5 4.3 1338 1 ACIN_MOUSE Q9JIX8 mus musculu
37 121.5 4.3 1783 1 RAA3_CHLRE Q9FEC4 chlamydomon
38 120.5 4.3 314 1 YK83_MYCTU Q10691 mycobacteri
39 120.5 4.3 1469 1 BCL9_DROME Q961D9 drosophila
40 120 4.3 706 1 HTF4_MOUSE Q61286 mus musculu
41 120 4.3 995 1 YIQ9_YEAST P40442 saccharomyc
42 120 4.3 1505 1 CUT1_HUMAN P39880 homo sapien
43 120 4.3 2541 1 TLN1_HUMAN Q9Y490 homo sapien
44 119 4.2 1953 1 BIGA_SALTY P25927 salmonella
45 119 4.2 2175 1 HMCU_DROME P10180 drosophila

Db 292 DVNEFDQVLPNPGHGVATHGQVTTTGYGSSASSAPAGAHMAKQPPQPPA 351
 Qy 269 TPDQAAANAESATKDLQTLQFAKNPENQKVNIDANGNAIPSGELXDDIVQIAQAKEA 328
 Db 352 QBPAAQ-----HTLPS-----TEREQQAQOR 372
 Qy 329 GVARQAQAVESNAQAQRYEDQHARRQBELQLSSGIG-----YGLSSALIVAGGIGAGV 382
 Db 373 PHIKTEQLSPHSNQOQHPEQOQQOQOOL-----CYGSPNLQHYGFSYPPI----- 419
 Qy 383 TIALHRNPQAEQTTHHTVVVQOQTGIGIPQHKVLMPEQRRRPSDRDSQGSVAS--T 440
 Db 420 -----TRSEYDTEHNSG-----SYSHAAGQSGSLYSTFT 451
 Qy 441 HWSDSSEVVNPYAEVGGARNLSAHOPEEH 471
 Db 452 YNNPQPMYPIADTSGVFTPIQTHSPQOH 482

RESULT 2

HLVA_SERMA STANDARD; PRT; 1608 AA.
 ID HLVA_SERMA STANDARD; PRT; 1608 AA.
 AC P15320;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 01-NOV-1990 (Rel. 16, Last annotation update)
 DE Hemolysin precursor.
 GN SHLA.
 OS Serratia marcescens.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Serratia.
 OX NCBI_TaxID=615;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 31-40.
 RC STRAIN=SN8;
 RX MEDLINE=88257037; PubMed=3290200;
 RA Poole K., Schiebel E., Braun V.;
 RT "Molecular characterization of the hemolysin determinant of Serratia
 marcescens";
 RL J. Bacteriol. 170:3177-3188(1988).
 CC -!- FUNCTION: BACTERIAL HEMOLYSINS ARE EXOTOXINS THAT ATTACK BLOOD
 CELL MEMBRANES AND CAUSE CELL RUPTURE BY MECHANISMS NOT CLEARLY
 DEFINED.
 CC -!- FUNCTION: CELL-BOUND HEMOLYSIN, WHICH RELEASES HEME-IRON FROM
 ERYTHROCYTES BY INTERACTION WITH THE ERYTHROCYTE MEMBRANE. SHLA
 REQUIRES SHLB FUNCTION.
 CC -!- SUBCELLULAR LOCATION: Outer membrane.
 CC -!- SIMILARITY: TO P.MIRABILIS HEMOLYSIN (HPMA).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; M22618; AAA50323.1; -
 DR PIR; A28182; A28182.
 KW Hemolysis; Toxin; Outer membrane; Signal.
 FT SIGNAL 1 30
 FT CHAIN 31 1608 HEMOLYSIN.
 SQ SEQUENCE 1608 AA; 165078 MW; D669B476FE7DAD51 CRC64;

Query Match 5.4%; Score 150.5; DB 1; Length 1608;
 Best Local Similarity 20.9%; Pred. No. 0.37;
 Matches 122; Conservative 84; Mismatches 236; Indels 141; Gaps 25;

Qy 33 GTCHLISSGALGRSLFSPLRNSMADSVDSRDPICGLPTNPSR-----LAAATSETC 84
 Db 1008 GSDLTVDKAGEGTOGRNSASSAQVTCGSDAAN--GINNVKDKALYQGTALNGRGKTA 1065

Qy 85 LLGGFEVLHDKGLDILNTOIGPSAFRVEVOA-----DQTH----- 120
 Db 1066 VNAAGGDIRLDOADKQSESR---SGFNVKASAKGGFTADSKNFGAGFGGTHNGESSST 1122
 Qy 121 AATGKNGLEVSTLSPQWSSLSQSIDTEGKNRFVFTGGRGSGHPMVTVASDIAEA-RT 179
 Db 1123 AQVNTISG-QQGVELKAGRLTLQGTDRKQGVLSAGN-----KVALQAEQTQT 1173
 Qy 180 RILAKLDPDNGHGRQPKVDTRSVGVGSASG-IDDGWVSETHSTTNSSVRSPKFWVS 238
 Db 1174 RKESKLSGNIDLGAGSSDSKEKTGNLSAGAFDIAKVNESATERQCATIASDGKVTLSA 1233
 Qy 239 ---GAIAGLAGLAATGIAQALALTPEDDPTTTPDQANAA---ESATKDLQTLQFAFK 292
 Db 1234 NGKGDALHLQGAQKVGSGSAAL-----EAKNGGILLBSAKNEQ----- 1271
 Qy 293 NPEN-----QKVNIDANGNAIP-----SGELXDDIVEQIAQAQAK-----E 327
 Db 1272 HKDNWSLGIKANAKGGQTFNKDAGGKVDPTNGKDTHTLGLAGLKVGVQEQQKTHANTGIT 1331
 Qy 328 AGEVARQAQAVESNAQAQRYEDQHARRQBELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
 Db 1332 AGDVTLSNGKDTRL-AGARVD-----ADSVQGVKGGDLHVESRKDVGVKQVDVDAGLS 1384
 Qy 388 RRNPQAEQTTHHTVVVQOQTGIGIPQ-----HKVALMPQER-----RRPSDRRDSQGS 436
 Db 1385 HSNDPGSSITSLSKSVGTPRYAGKVKLEAGVKNKADATTDKYNVSARRLDPOQDTTGA 1444
 Qy 437 VASTHMSDSSEVVNPYAEVGGARNLSAHOPEEHIVDEVAADPGYVINFSGSGPVTG 496
 Db 1445 VS---FSKAGKVTLPATPAG-----EKQPLMDRGAIVTGVGAVKSDITGPAGRQG 1493
 Qy 497 RL-----IGTPGQGIQSTVALLANSGLRLGMGLT---SGGE 531
 Db 1494 HLKVNADVNNNAVGEQSA---IAGKNGVALQVGGQTLTGGE 1533

RESULT 3

Y456 CHLTR STANDARD; PRT; 1005 AA.
 AC O84452;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Protein Y456 precursor.
 GN CT456.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/Cx;
 RX MEDLINE=93000809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 Davis R.W.;
 RA "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis";
 RT Science 282:754-759(1998).
 RL -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL CPN0572/CT456/TC0741
 CC FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC EMBL; AE001319; AAC68056.1; -
 DR PIR; C71513; C71513.
 DR PHCI-2DPAGE; O84462; -.

CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
 CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
 CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
 CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
 CC AND G3.
 CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
 CC CHAINS. N-LINKED AND O-LINKED OLIGOSACCHARIDES.
 CC -1- DISEASE: DEFECTS IN AGC1 ARE THE CAUSE OF CARTILAGE MATRIX
 CC DEFICIENCY (CMD), AN AUTOSOMAL RECESSIVE SYNDROME CHARACTERIZED BY
 CC CLEFT PALATE, SHORT LIMBS, TAIL AND SNOUT. MUTATION IN STRAIN CMD
 CC CAUSES ABSENCE OF AGGREGAN BY TRUNCATION OF THE PROTEIN (MUTATION
 CC IN THE G1 DOMAIN).
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
 CC -1- SIMILARITY: Contains 4 link domains.
 CC -1- SIMILARITY: Contains 1 EGF-like domain.
 CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
 CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
 CC -1- SIMILARITY: BELONGS TO THE AGGREGAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L07049; AAC37670.1; -;
 CC EMBL; S73722; AAB32160.1; -;
 CC EMBL; S73721; AAB32160.1; JOINED.
 CC PIR; A55182; A55182.
 CC HSP; P98066; 1TSG.
 CC MGD; MGI:99602; Agcl.
 CC InterPro; IPR002353; Antifreeze1.
 CC InterPro; IPR007110; Ig-like.
 CC InterPro; IPR003006; Ig MHC.
 CC InterPro; IPR003596; Ig v.
 CC InterPro; IPR001304; Lectin_C.
 CC InterPro; IPR000538; Link.
 CC InterPro; IPR003324; SGXSG.
 CC InterPro; IPR000436; Sushi_SCR_CCP.
 CC Pfam; PF00047; Ig; 1.
 CC Pfam; PF00059; lectin_c; 1.
 CC Pfam; PF02339; SGXSG; 60.
 CC Pfam; PF00084; sushi; 1.
 CC Pfam; PF00193; Xlink; 4.
 CC PRINTS; PRO1285; LINKMODULE.
 CC PRINTS; PRO0356; ANTIFREEZEII.
 CC ProDom; PD000918; Link; 4.
 CC SMART; SM00032; CCP; 1.
 CC SMART; SM00034; CLECT; 1.
 CC SMART; SM00406; IGV; 1.
 CC SMART; SM00445; LINK; 4.
 CC PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 CC PROSITE; PS50041; C-TYPE LECTIN_2; 1.
 CC PROSITE; PS50835; IG LIKE; 1.
 CC PROSITE; PS00290; IG MHC; 1.
 CC PROSITE; PS01241; LINK; 4.
 KW Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
 KW Repeat; Immunoglobulin domain.
 FT SIGNAL 1 19 POTENTIAL.
 FT CHAIN 20 2132 AGGREGAN CORE PROTEIN.
 FT DOMAIN 34 147 IG-LIKE V-TYPE.
 FT DOMAIN 170 247 LINK 1.
 FT DOMAIN 268 349 LINK 2.
 FT DOMAIN 504 581 LINK 3.
 FT DOMAIN 602 683 LINK 4.
 FT DOMAIN 1918 2044 C-TYPE LECTIN.
 FT DOMAIN 2048 2106 SUSHI.
 FT DOMAIN 48 140 GI-A.
 FT DOMAIN 152 247 GI-B.
 FT DOMAIN 253 349 GI-B'.
 FT DOMAIN 486 580 G2-B.

FT	DOMAIN	587	682	G2-B'.
FT	DOMAIN	685	803	KS.
FT	DOMAIN	805	1231	CS-1.
FT	DOMAIN	1232	1917	CS-2.
FT	DOMAIN	1917	2132	G3.
FT	DISULFID	51	133	BY SIMILARITY.
FT	DISULFID	175	246	BY SIMILARITY.
FT	DISULFID	199	220	BY SIMILARITY.
FT	DISULFID	273	348	BY SIMILARITY.
FT	DISULFID	297	318	BY SIMILARITY.
FT	DISULFID	509	580	BY SIMILARITY.
FT	DISULFID	533	554	BY SIMILARITY.
FT	DISULFID	607	682	BY SIMILARITY.
FT	DISULFID	631	652	BY SIMILARITY.
FT	DISULFID	1922	1933	BY SIMILARITY.
FT	DISULFID	1950	2042	BY SIMILARITY.
FT	DISULFID	2018	2034	BY SIMILARITY.
FT	DISULFID	2049	2092	BY SIMILARITY.
FT	DISULFID	2078	2105	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	239	239	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	333	333	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	387	387	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	611	611	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	667	667	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	CARBOHYD	1675	1675	N-LINKED (GLCNAC. .) (POTENTIAL).
FT	SITE	1171	1173	CELL ATTACHMENT SITE (POTENTIAL).
SQ	SEQUENCE	2132	AA; 222008	MM; 0B2BCDFC6CDBA163 CRC64;

Query Match 5.0%; Score 140; DB 1; Length 2132;
 Best Local Similarity 21.0%; Pred. No. 2.2;
 Matches 130; Conservative 82; Mismatches 249; Indels 158; Gaps 27;

Qy	4	GNLGNVNGNHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFPLNSMADSVDS	63
Db	906	GDLGLPSGGEGITETSGAEETSGLPSSGGG-LETSTSGVDDVSGIPTGREGLETSSAG	964
Qy	64	-RDIPGLPTNPRLAAATSE---TCLLGGPEVLHD--KGLDILINTQIGPSAFREYVQA	116
Db	965	VEDLSGLPSGGEGSETSTSGIEDISVLPTGSELETSSAGVGLSLGSPGGES--LETS	1022
Qy	117	DGTHAAI---GEKNGLVS-----VTLSPQWMSLQSIDTEGKNRFFVTGGRGSGHPM	167
Db	1023	SGAEDVTQLPTEGGLTSSAGVEDITVLPTGRESLETSSAGVEDVGLSPSGREG---L	1078
Qy	168	VTVASDIAEARTRILAKLDPNHGGRQPKVDVTSVGVGSASGIDGVDVSTHSTTNS	227
Db	1079	ETSASGIED-----ISVFPTAEAG-----LDTASAG-GYVSGIPSG--GDGTET	1123
Qy	228	VRSDPKFWVSYGAIAGAAGL--AATGIAQALALTPEPDDPTTTPDQAAANAESATKDQ	285
Db	1124	VE-----DVSGLPSSGGEGLETSSAGV-----EDLGPSTRDSLETSSAGVDVT	1165
Qy	286	LTOEAFK-----NPNQKNVINDANGNAIPSGELXDDIVEQIAQAQAKEAGEVARQA	341
Db	1166	----GFPSSGRGDPETSVSGVGDGDFGLPSG-----KEGLETS	1199
Qy	342	QAQRYEDQHARQBEELQSSG-----	366
Db	1200	SGAEDLSGLPSGKEDLVGSASGALDFGKLPPFTLGGTQTPVNGPSPGSGEYSAGD	1259
Qy	367	GLSSALI VAGGIGAGVTTALHRRNQPAEQTTTTTHTVVQQTGGIP-QHKVALMPQ	425
Db	1260	GPSSGLPDPFGLPSGFPPTVSLVDSTLVITATTTSSELEGRGTIGISSGSGVGLP	1317
Qy	426	RFSRRDSQGSVASTHMSDSSEVNVNVAEYGGARNSLSAHPQBEHIYDEVAADP--	483
Db	1318	ELDSSADISGLPSGTSLSGQASGSPDSGSETSG-----FFDVSGQPPGSSG	1363
Qy	484	VIONPSSGGPVTGRILGTPPGGIGISTVALL-----ANSGGLRLGNG---GLTS-	528
Db	1364	VSEETSGIPEISGQPSGTPDTTATTSVGTVELNELSSGQPDVSGDGSGLIFGSGSSG	1423

RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivaundaram S.,
RA Tagami H., Takahashi H., Takeda J., Takenoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
CC -1- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -1- SIMILARITY: STRONG, TO BORDETTELLA PERTACTIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U36840; AAA79815.1; ALT SEQ.
CC EMBL; AE000350; AAC75695.1; -
CC EMBL; D90889; BAA16514.1; ALT INIT.
CC EMBL; D90890; BAA16518.1; ALT_INIT.
CC PIR; A65044; A65044.
CC EcoGene; EG13213; YpJA.
CC InterPro; IPR006315; Autotransporter.
CC InterPro; IPR005546; Autotransporter.
CC InterPro; IPR004899; Pertactin.
CC InterPro; IPR003991; Pertactin_C.
CC Pfam; PF03797; Autotransporter; 1.
CC Pfam; PF03212; Pertactin; 1.
CC PRINTS; PR01484; PERTACTIN_FAMILY.
CC TIGRFAMs; TIGR01414; autotrans_bar1; 2.
CC Hypothetical protein; Outer membrane. Complete proteome.
KW Hypothetical protein; Outer membrane.
SQ SEQUENCE 1569 AA; 162774 MW; 0D06A9A32154596 CRC64;

Query Match 4.9%; Score 136; DB 1; Length 1569;
Best Local Similarity 20.9%; Pred. No. 2.5;
Matches 127; Conservative 69; Mismatches 259; Indels 154; Gaps 26;

QY 19 APPLPSQDTGAARGT--GHLISSTGALGSRSLFPLNSMADSVDSRDPLGPTNPRL 76
DB 579 AKHVQSQGGALIASTTSGTLIEGNSYGFAPY---IRNSEAKV-----VLENSAGSL 628
QY 77 AAATSETCLGGFEVLHDKGPDILNTOIGP---SAFRVEVQADGTHAAIGKNGLEVSU 133
DB 629 TVVTGSRV---DTIANGKMDVTKGVTVNLNAGTQTIYASATSKANIKGKQIVY 685
QY 134 TLSPOE--WSSLSQSDTEGKRFVFTGGRGGSGHPMTVVASDIABARTILAKLPD--- 188
DB 686 GLATEANIESGRQIVDGGSTEXTHINGGTQTVQNYGKAIINTDIVSGLQIIMANGTAEGSI 745
QY 189 -----NHGRQPKDV-----DTRSVGVGSAGI---DDGVVSETHSTTNSVRS 230
DB 746 INGSQVWNEGGLAENSVLNDGTTLDVRE--KGSATGIQQSSQSGALVATTRATRVGTGA 803
QY 231 DPKFWVSGAIAAGLAGLAATGCI--AQALALPEPDPTTTPDQAAANAESATKDQLTQ 288
DB 804 D-----GVAFSEQGGANWILLANGVLIVESD--TSSDKTVQNMGGREIVKAT- 852
QY 289 EAFKNPENQKVNIDANGNAIPSGELXDDIVEQIA-----QQAKEAGEVARQQ--- 335
DB 853 -----ATGTTLTGGE--QIVEGVANETTINDGGITQTVSANGAIAIKTINE 895
QY 336 -----AVESNAQOQRYEDQHARRQE---LQLSSGIGVGLSS-----A 371
DB 896 GGTLLTVNDGKATDIVQNSGAALQTSANGIBISGTHQVGFSTISGNLATNMLENGNL 955
QY 372 LIVAGGIGAVGVTALHRRNQPAEQTTTTHTVVQOQTGGIPQHKVLMQPERRFSRDR 431
DB 956 LVLVLAETARDSTVGVKGAMQNLGQDSATKVN-----SGG--QYTLGRSKDBFQALARE 1007
QY 432 DSQ--GSVASTHSDSSSEVVNPNVYAEVGGARNLSAHPQEEHIYDEVAADPGYSVIQNF 489

DB 1008 DLQVAGGTAIYVAGTLDAD-----ASVSGATGSLMTPTPRDNV-----TPVKLE 1050
QY 490 GSGPVTRGLIGTGGQIGSTVALL--ANSGLRLGMLGTLTSG-----GET 532
DB 1051 GAVRITDSATITLNGVDTTLDADLTAAARGSVLNSNNSCAGTSNCEYRVNSLLNDGDV 1110
QY 533 AVSSVNAAP 541
DB 1111 YLSAQTAAP 1119

RESULT 8
PCGA RAT
ID PGCA RAT
AC P07897; STANDARD; PRT; 2124 AA.
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE protein) (CSPCP).
GN AGC1 OR AGC.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88087070; PubMed=3693370;
RA Doerge K., Sasaki M., Horgan E., Hassell J.R., Yamada Y.;
RT "Complete primary structure of the rat cartilage proteoglycan core
RT protein deduced from cDNA clones.";
RL J. Biol. Chem. 262:17757-17767(1987).
RN [2]
RP REVISION TO 698.
RA Doerge K., Sasaki M., Horgan E., Hassell J.R., Yamada Y.;
RL J. Biol. Chem. 263:10040-10040(1988).
RN [3]
RP SEQUENCE OF 1856-2124 FROM N.A.
RX MEDLINE=86250698; PubMed=2424893;
RA Doerge K., Fernandez P., Hassell J.R., Sasaki M., Yamada Y.;
RT "Partial cDNA sequence encoding a globular domain at the C terminus
RT of the rat cartilage proteoglycan.";
RL J. Biol. Chem. 261:8108-8111(1986).
CC -1- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC MATRIX OF CARTILAGINOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC -1- SUBUNIT: Interacts with FBLN1 (By similarity).
CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC similarity).
CC -1- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC AND G3.
CC -1- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -1- SIMILARITY: Contains 4 link domains.
CC -1- SIMILARITY: Contains 1 C-type lectin family domain.
CC -1- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC

glycosylated domain.",
J. Biol. Chem. 266:9678-9686 (1991).
[2]
SEQUENCE OF 1-503 FROM N.A.
TISSUE=Submaxillary gland;
MEDLINE=88087170; PubMed=2826455;
TAMPE C.S., Eckhardt A.E., Abernethy J.L., Hill R.L.;
"Porcine submaxillary gland apomucin contains tandemly repeated,
identical sequences of 81 residues."
J. Biol. Chem. 263:1081-1088 (1988).
[3]
SEQUENCE OF 45-80.
TISSUE=Submaxillary gland;
MEDLINE=87280230; PubMed=3611111;
Eckhardt A.E., Timpie C.S., Abernethy J.L., Toumadje A.,
Johnson W.C. Jr., Hill R.L.;
RT "Structural properties of porcine submaxillary gland apomucin.";
J. Biol. Chem. 262:11339-11344 (1987).
[4]
CARBOHYDRATE-LINKAGE SITES, AND SEQUENCE OF 45-125.
TISSUE=Submaxillary gland;
MEDLINE=97248516; PubMed=9092502;
Gerken T.A., Owens C.L., Pasumathary M.;
RT "Determination of the site-specific O-glycosylation pattern of the
porcine submaxillary mucin tandem repeat glycopeptide. Model proposed
for the polypeptide:galnac transferase peptide binding site.";
J. Biol. Chem. 272:9709-9719 (1997).
CC -1- FUNCTION: APOMUCIN IS PART OF MUCIN, THE MAJOR GLYCOPROTEIN
SYNTHESIZED AND SECRETED BY MUOUS CELLS OF THE SUBMAXILLARY
GLAND. ITS HIGHLY VISCOUS AQUEOUS SOLUTIONS SERVE TO LUBRICATE
THE ORAL CAVITY AND TO PROTECT IT FROM THE EXTERNAL
ENVIRONMENT.
CC
CC -1- SUBUNIT: INTERMOLECULAR DISULFIDE BONDS COULD HELP MAINTAIN A
MULTIMERIC MUCIN STRUCTURE.
CC
CC -1- SUBCELLULAR LOCATION: Secreted.
CC
CC -1- TISSUE SPECIFICITY: SUBMAXILLARY MUCOSAE.
CC
CC -1- DOMAIN: CONTAINS TANDEMLY REPEATED, IDENTICAL SEQUENCES OF 81
RESIDUES.
CC
CC -1- PTM: EXTENSIVELY O-LINKED GLYCOSYLATED ON SER AND THR RESIDUES OF
THE REPEAT UNITS. HIGHEST GLYCOSYLATION APPEARS TO OCCUR ON SER
RESIDUES WHICH HAVE GLY AT POSITIONS AT +2 OR -2 FROM THE
GLYCOSYLATION SITE OR, WHERE GLY IS THE PENULTIMATE RESIDUE. THE
PRESENCE OF PROLINE (USUALLY AT POSITION +3 OR -3) APPEARS TO ALSO
ENHANCE GLYCOSYLATION.
CC
CC -1- SIMILARITY: Contains 1 VWFC domain.
CC
CC -1- SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; M61883; AAA30998.1; --
EMBL; M21174; AAA30990.1; --
DR InterPro; IPR006208; Cys_knot.
DR InterPro; IPR006207; Cys_knot_C.
DR InterPro; IPR001007; VWF_C.
DR Pfam; PF00007; Cys_knot; 1.
DR SMART; SM00214; VWG; 1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01235; CTCK_2; 1.
DR PROSITE; PS01208; VWFC_1; 1.
DR PROSITE; PS01084; VWFC_2; 1.
KW Glycoprotein; Repeat.
FT NON_TER 1
FT DOMAN <1 368 81 AA TANDEM REPEATS.
FT REPEAT <1 44 1.
FT REPEAT 45 125 2.
FT REPEAT 126 206 3.
FT REPEAT 207 287 4.

```

Db 370 SVAGSGGAPVSSG-----ASQA--AGTSEATTSIEGAG----- 401
Qy 358 LQLSSGIGVGLSALLIVAG-----GICAGVTTALHRRN-OP-----AEQITTTT-----TT 401
Db 402 ---TSGVGF-KTEATTPGNETTRVGTATGTTGIVSKTLPEGYNTEATTSIGRSCTT 457
Qy 402 HTVQQOQTGGIPQHKVALMPQRRRFSRRRDSQGSVASTHSDSSSEVVNPPVAEYGGARN 461
Db 458 HT-----DLPGGTTIVLPGFSH--SSQSSKPGSSVTPGSPESGSETGTSGEPSTTVI 508
Qy 462 SLUSAHOPEHIYDEVAADPGYIVQNFSGSGP-VTGRLLICP-GGIGIOSTVALLANSGL 519
Db 509 SGSSH-TEATTPIGGSGSP-----TGSREGTGTCGELSGTTIASGNATTEA--TTSTET 558
Qy 520 RIG-----MGGLTSGTAVSVNAAP 541
Db 559 RIGPQTGAQTTVPQSGVSGSETGTSEAVSNP 589

RESULT 10
SLAP CAUCR
ID SLAP CAUCR STANDARD; PRT; 1025 AA.
AC P35828; Q46015; Q9RF12;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE S-layer protein (Paracrystalline surface layer protein).
GN RSAA OR CC1007.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteraceae; Caulobacter.
OX NCBI_TaxID=155892;
RN
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-21; 646-660 AND 1021-1025.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=93007489; PubMed=1393820;
RA Gilchrist A., Fisher J.A., Smit J.K.;
RT "Nucleotide sequence analysis of the gene encoding the Caulobacter
rescentus paracrystalline surface layer protein.";
RL Can. J. Microbiol. 38:193-202(1992).
RN
[2]
RP REVISIONS TO 376; 636 AND 842-843.
RA Awram P.;
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=JS3001;
RA Bingle W.H., Awram P.A., Nomellini J.F., Smit J.K.;
RT "The secretion signal of C. crescentus S-layer protein is located in
the C-terminal 82 amino acids of the molecule.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
RN
[4]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heideberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Porocka I., Nelson W.C., Newton A.S., Stephens C., Phadke N.D., Ely B.,
RA DeBooy R.T., Dodson R.J., Durkin A.S., Gwin M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Uterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN
[5]
RP SEQUENCE OF 1-313 FROM N.A., AND SEQUENCE OF 1-20.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=89008089; PubMed=3049545;
RA Fisher J.A., Smit J.K., Agabian N.;
RT "Transcriptional analysis of the major surface array gene of
Caulobacter crescentus.";
RL J. Bacteriol. 170:4706-4713(1988).

```

```

RN [6]
RP CHARACTERIZATION.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=98292737; PubMed=9620954;
RA Awram P., Smit J.K.;
RT "The Caulobacter crescentus paracrystalline S-layer protein is
secreted by an ABC transporter (type I) secretion apparatus.";
RL J. Bacteriol. 180:3062-3069(1998).
CC -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY
OF PROTEINS WHICH COAT THE SURFACE OF BACTERIA. PROBABLY ACTS AS A
SUBCELLULAR LOCATION: CELL WALL. THIS BACTERIUM IS COVERED BY A S-
LAYER WITH HEXAGONAL SYMMETRY. SECRETED BY AN ABC TRANSPORTER
(TYPE I) SECRETION APPARATUS.
CC -!- MISCELLANEOUS: THE CALCIUM BINDING OF THIS PROTEIN MAY MEDIATE
SURFACE ATTACHMENT BY IONIC BRIDGING BETWEEN THE S-LAYER AND A
SPECIFIC MEMBRANE-ASSOCIATED OLIGOSACCHARIDE.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
DR EMBL; AF062345; AAC38665.2; -
DR EMBL; AF193063; AAF19365.1; -
DR EMBL; AE005779; AAK22991.1; ALT_INIT.
DR PIR; A48995; A48995.
DR HSP; P22629; ISWC.
DR TIGR; CC1007; -
DR InterPro; IPR001343; Hemlysn_Ca_bind.
DR Pfam; PF00353; hemolysinCabin; 3.
DR PRINTS; PR00313; CABNDNGRPT.
KW Cell wall; S-layer; Calcium-binding; Complete proteome.
FT INIT_MET 0
SQ SEQUENCE 1025 AA; 98001 MW; AD7A326E1363D8AC CRC64;

Query Match 4.7%; Score 132.5; DB 1; Length 1025;
Best Local Similarity 21.0%; Pred. No. 2.3;
Matches 117; Conservative 66; Mismatches 232; Indels 143; Gaps 24;

Qy 26 TDGAARGTGHLSSTGALGSRSLFSPLRNSMADSVSRDIPGLTPSPRLAATSETCL 85
Db 229 TDNAAGVNLFTAYFPSSGVSGSTL-----SLATGTDT--LTGTANNDFVAGEVAGAA 279

Qy 86 LGGEVLHDKGLDILN-----TQIGPSAFRVEVQDGTTHAAIGKNGLEVSVT 134
Db 280 LTVGDTLGGAGTDLVNWQAAVATLPTGVTISGIETMNTVSGAATLNTSSGVTLTA 339

Qy 135 LSPQEWSSLSQSIDTEGKNRFVFTGGRGSGHPMTVVASDIAEARTRILAKLDPNHGRQ 194
Db 340 LNTWTSGRAQIVTAGAGQNLATTAQAANNVAVDGGANVTAVSTGVTS----- 388

Qy 195 PROVDTRSVGVSGASGIDGVVSETHSTTNSVSRDPKFWVSVGAIAAGLAGLAATGIA 254
Db 389 ----GTTTVGANSAA---SGTVS---SVANSST-----TTTGAIA--VTGGTAVTVA 429

Qy 255 QALALTPEPDDPTTDDPQANAAESATKDQLTQEAQKNPENQKNVIDANGNAIPSGELX 314
Db 430 QTAG---NAVNTLTQADVTVTGNSSSTTAVTVTQTA-----AATGATVAGRVN 475

Qy 315 D--DIVEGIAQQAQKEAGEVAR-----QOAVESNAQAQRYEDQHARRQEBELQLSSGIG 365
Db 476 GAVTITDSAAASATTAGKIATVTLGSGFAATIDSSALTTVNL-----SGTG 521

Qy 366 YGLSALLIVAGIGAGVTTALHRRNQPAEQTTTTTHTVVOQTGGIPQHKVALMPQERR 425
Db 522 TSL-----CIGRGALTA-----TPTANTLTLLNVNGLT--TTGATIDSEAA----- 559

Qy 426 RPSDRRDSQGSVASTHSDSSSEVVNPPVAEV---GGARNLSAHOPEHIYDEVAADPG 481

```

Db 560 --ADDGTTTINIAGSTASTIASLVADATTINISGDARVTITSH-----TAAALTG 609

Qy 482 YSVIQNFS--GSGEVTGRLGTGPGIQS-----TYALLANGS-----GRLGMGG 525

Db 610 ITVTNVTGATLGAELATG-LVFTOGAGADSILLGATTKAIVMGAGDDTIVTSSATLGAGG 668

Qy 526 LTSGS---ETAVSSVNA 540

Db 669 SVNGGDDGTDLVANVNGS 686

RESULT 11

FLID_VIBCH STANDARD; PRT; 666 AA.

AC Q9K063;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Flagellar hook-associated protein 2 (HAP2) (Filament cap protein)

DE (Flagellar cap protein)

GN FLID OR VC2140.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;

OC Vibrionaceae; Vibrio.

OX NCBI_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=El Tor N16961 / Serotype O1;

RX MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Unayam L.A.,

RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,

RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,

RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.W.;

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

RT cholerae.";

RL Nature 406:477-483(2000).

CC -!- FUNCTION: REQUIRED FOR THE MORPHOGENESIS AND FOR THE ELONGATION OF

CC THE FLAGELLAR FILAMENT BY FACILITATING POLYMERIZATION OF THE

CC FLAGELLIN MONOMERS AT THE TIP OF GROWING FILAMENT. FORMS A CAPPING

CC STRUCTURE, WHICH PREVENTS FLAGELLIN SUBUNITS (TRANSPORTED THROUGH

CC THE CENTRAL CHANNEL OF THE FLAGELLUM) FROM LEAKING OUT WITHOUT

CC POLYMERIZATION AT THE DISTAL END (BY SIMILARITY).

CC -!- SUBUNIT: Homopentamer (By similarity).

CC -!- SUBCELLULAR LOCATION: Flagellar.

CC -!- SIMILARITY: BELONGS TO THE FLID FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AE004287; AAF95285.1; -.

DR PIR; G82111; G82111.

DR TIGR; VC2140; -.

DR InterPro; IPR003481; Flid.

DR Pfam; PF02465; Flid; 1.

KW Flagella; Coiled coil; Complete proteome.

FT DOMAIN 354 419 COILED COIL (POTENTIAL).

SQ SEQUENCE 666 AA; 72330 MW; 6794EFEC34A2A0D2 CRC64;

Query Match

Best Local Similarity 22.4%; Score 132; DB 1; Length 666;

Matches 148; Conservative 91; Mismatches 249; Indels 174; Gaps 37;

Qy 2 PIG-NLGNVNGN-HLIPAPPLPSQTD-GAARGTGTHLSSTGALGS-----RSLFSP 53

Db 5 PMGNTGPDINGMWSKIVSAERVPKQQRIDNERTNIDSISAYGRLSRLDTMKNLMTQF 64

Qy 54 RNSWADSVDSRDIFGLPTNPSRLAAATSETCLLG-----FEVLHDK-- 95

Db 65 ROEKAFARVVD-----TSNEQVVSATATTETAIAGNTSVDVLQLAQSHKIASEVL-DKDA 118

Qy 96 --GPLDILNTQIGPSARFVEVQADGTHAAI-----GEKN--GLEVSVTLSPOEWSLSQSI 146

Db 119 KFGP-GKLHISLGDKSFSLDVQGNKSLVDIVRGINGEKSPGVRAISINDVEGRLIVAS 177

Qy 147 DTEGKRFVFTGGGGGHPMTVTVASDIAEARTR-----ILAKLPDNDHGGQP 195

Db 178 NVSGKHSVMSAQAEFGNPLKQLEYKTLQORVRLDEKARAQAOQLIAPLTP-----QQ 232

Qy 196 KVDTRSVGVSGASGIDGVVSVTHSTTNSSVSRDPKFWVSVGAIAAGLAGLAATGIAQ 255

Db 233 KVAAKVAEKIGDAARLVQDEVAQIRSAQS-----AQGAAGEALNA--- 274

Qy 256 ALALTPEPDDPTTDDPQANAAESATK-----DQL-----TOEAFKNPENQKN 300

Db 275 -----GELTSANVKAANAASEAKKYIRPDIIPGWTTETASGTLTLDYSWEPEE--- 323

Qy 301 IDANGNA-----IP-----SGELXDD-IVEQIAQOAKE---AGEVARQAV----- 337

Db 324 LDAQGQKKAADVPCWSNTASGTLTLDYSVTPQEAQOKLEQKLAQEAQIAEAIIRSGKMTPE 383

Qy 338 ESNAQA-----QORYEDQHARQEELQLSSGI--GYG-----LSSALIVAGGIGAG 381

Db 384 EAKAQARAKLSPERAVIEQVEKAQAALNAASAFDGYGGMTEVQSAQDSMVLDGV--- 440

Qy 382 VTTALHRRNPQAEQTTHHTVVVQQTGGIPQHKVALMPQERRRFRSRRRSQSVAS-T 440

Db 441 --ATLSSNNNIENAEIGVNLTKGKTRDNPPEAIEGI---EYDRVRNDIEQFVAAYN 495

Qy 441 HWSDSSEV--VNPVAEYGG--ARNSL--SAHQPEEHYDEVAADPGYSVITQNFSGSPV 494

Db 496 QFFTSKELACVDPTQAGPLAGDSIVRSADSLKIVFSS-----SIQAPENLKS 548

Qy 495 TGRIGTGGQG-IQSTYALL---ANSGLRL---GMGLTSGGETAVSSVNAAPTGP 545

Db 549 TFGITTTQCTLEINVMALDRLNNFTKLGEFFGNGQFQAKRVEDAISMTG--VTGS 606

Qy 546 VR 547

Db 607 IR 608

RESULT 12

CSG_HALVO STANDARD; PRT; 827 AA.

ID - CSG HALVO

AC P25062;

DT 01-MAY-1992 (Rel. 22, Created)

DT 01-MAY-1992 (Rel. 22, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Cell surface glycoprotein precursor (S-layer protein).

OS Halobacterium volcanii; Halobacter volcanii).

OC Archaea; Euryarchaeota; Halobacteria; Halobacteriales;

OC Halobacteriaceae; Halobacter.

OX NCBI_TaxID=2246;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RX MEDLINE=91072265; PubMed=2123862;

RA Sumper M., Berg E., Mengele R., Strobel I.;

RT "Primary structure and glycosylation of the S-layer protein of

RT Halobacter volcanii";

RL J. Bacteriol. 172:7111-7118(1990).

CC -!- FUNCTION: THE S-LAYER IS A PARACRYSTALLINE MONO-LAYERED ASSEMBLY

CC OF PROTEINS WHICH COAT THE SURFACE OF THE CELL.

CC -!- SUBCELLULAR LOCATION: Cell wall. This archaea is covered by an S-

CC layer with hexagonal symmetry.

CC -!- PTM: PROTEIN IS CONTAINS O-LINKED GLYCANS WHICH CONSIST OF GLC-GAL

CC DISACCHARIDES.

CC -!- SIMILARITY: TO H.HALOBUTUM AND H.JAPONICA CELL SURFACE

CC GLYCOPROTEINS.


```
FT DOMAIN 738 743 POLY-THR.
FT DOMAIN 774 779 POLY-THR.
FT DOMAIN 874 877 POLY-SER.
FT CARBOHYD 471 471 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 579 579 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 687 687 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 723 723 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 820 820 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 886 886 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 918 918 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 973 973 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1045 1045 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1068 1068 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 1260 AA; 132641 MW; 763D1063A2354C24 CRC64;

Query Match 4.6%; Score 130; DB 1; Length 1260;
Best Local Similarity 20.6%; Pred. No. 4.2;
Matches 73; Conservative 48; Mismatches 167; Indels 66; Gaps 12;

Qy 16 IPPAPPLP---SQTGGAARGGTGHLISS-----TGALGSRSLFSPLRNSMADSVDSRD 65
Db 895 VPTASTMSDLSSTDGISATSSDNDVSKGVSVTTEISVTTIQTTPNPLSSSVTLTQLSS 954

Qy 66 IPGLPTNPRLAAATSETCLLGGFEVLHDKGFLDILNTQIGPSAFRV--EVQADGTHAAI 123
Db 955 IPSVSESESKVFTTSNGDNQSTHDSQSTSTEIEIVTT-----SSTKVLPPVVSNTDLTS 1010

Qy 124 GEKNGLEVSVTLSPQWMSLOSIDPE---GKNRFVFTGGRGGSGHPMTVA-SDIAEART 179
Db 1011 EPTNTRQOPTTLSTTSNSITDITTSQPTGDN-----GDNTSSTNPVPTVATSTLASAS- 1064

Qy 180 RILAKLDPDNHCGRPKDVRFSVGVSGASGIDDVSVSETHSTTNSVRSRDPKFWVSVG 239
Db 1065 -----EEDNKGSH-----ESASTSLKPSMGNSGLTSTTEIATTSPTTEAPSPAVSSG 1114

Qy 240 AIAAGLAGLATGIAQALATPEPD-----PTTDPQAAANAASATKDQLTQEAFAKNPE 295
Db 1115 T-----DVTTEPTDTRQOPTTLSTTSKTNSSESVATTQATNENCGKSPS 1157

Qy 296 NQ-----KVNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQAAVESN 340
Db 1158 TDLTSLATGTSASTSANSSELVTSVTCGAVASANDQSHSTSVTNSNSIVSN 1211
```

Search completed: December 16, 2003, 08:53:50
Job time : 24.306 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 26.2608 Seconds
(without alignments)
2010.470 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNVNGNHLPPAP.....GETAVSSVNAAPTGPVRV 549
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616862 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1528.5	54.6	558	2 A98199	translocated intim
2	1528.5	54.6	558	2 E85045	probable transloca
3	151.5	5.4	2793	2 B90784	hypothetical prote
4	150.5	5.4	1608	2 A28182	hemolysin A - Serr
5	150	5.4	1005	2 C71513	hypothetical prote
6	149.5	5.3	2806	2 D85644	hypothetical prote
7	149	5.3	2232	2 T34434	hypothetical prote
8	146	5.2	1829	2 T24583	collar protein iso
9	144.5	5.2	1296	2 T13936	puff 748 protein -
10	144	5.1	883	2 S04722	submaxillary mucin
11	144	5.1	1589	2 T42233	hypothetical prote
12	143.5	5.1	2551	2 B98047	hypothetical prote
13	141.5	5.1	647	2 T39141	hypothetical prote
14	141	5.0	2335	2 AC0304	probable hemolysin
15	140	5.0	1152	2 AE1852	hypothetical prote
16	140	5.0	2132	1 A55182	aggreccan precursor
17	137	4.9	1532	2 A26039	IGA-specific metal
18	137	4.9	2468	2 A83412	hypothetical prote
19	136.5	4.9	770	2 T51024	related to C2H2 zi
20	136.5	4.9	1829	2 T35681	probable sensory h
21	136	4.9	674	2 B82117	probable flagellar
22	136	4.9	1569	2 A65044	hypothetical prote
23	135.5	4.8	2124	2 A28452	proteoglycan core
24	135	4.8	1085	2 S66149	gene pipsneak pro
25	134.5	4.8	3295	2 AE0074	probable adhesin Y
26	134.5	4.8	13288	2 T03099	mucin, submaxillar
27	134	4.8	3013	2 AB0480	probable invasini
28	133.5	4.8	1975	2 B81192	hemagglutinin/hemo
29	133.5	4.8	1995	2 G81044	hemagglutinin/hemo

hypothetical glyci
hypothetical prote
hemagglutinin-like
hemagglutinin-like
probable zinc-fing
paracrystalline su
S-layer protein Rs
ascites sialoglyco
flagellar hook-ass
S-layer protein -
hypothetical prote
probable RTX famil
hypothetical prote
hypothetical prote
hypothetical prote
notch protein - fr

30 133.5 4.8 2174 2 B95965
31 133 4.8 1571 2 C91068
32 133 4.8 3442 2 B82589
33 133 4.8 3455 2 B82519
34 132.5 4.7 812 2 T48016
35 132.5 4.7 1026 2 A48995
36 132.5 4.7 1073 2 C87374
37 132.5 4.7 1630 2 A53577
38 132 4.7 666 2 G82111
39 132 4.7 827 2 A37849
40 132 4.7 1528 2 D85912
41 132 4.7 5188 2 B85547
42 132 4.7 5291 2 F90696
43 131.5 4.7 635 2 F75477
44 131.5 4.7 1804 2 H96597
45 131 4.7 2703 1 A24420

ALIGNMENTS

RESULT 1
A98199
translocated intimin receptor Tir [imported] - Escherichia coli (strain O157:H7, substra
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: A98199
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: A98199
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-558 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA837984.1; PID:g13364036; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: EC84561

Query Match 54.6%; Score 1528.5; DB 2; Length 558;
Best Local Similarity 56.2%; Pred. No. 6.5e-88;
Matches 327; Conservative 60; Mismatches 138; Indels 57; Gaps 11;

Qy 1 MPIGNLGNVNGNHLIPAPLPQSOTGAAGGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNVNGNHLIPAPLPQSOTGA--GGRGQLINSTGPGSRALFTPVNSMADS 58
Qy 61 VDSR--DIPGLFTNPRLAAATSETCLGGFEVLHDKGLDILNTQIGPSAFRVEVQADG 118
Db 59 GDNRASDVPLPVNPMRLAA--SEITLNDGFEVLHDKGLDILNTQIGSVFRVETQEDG 116
Qy 119 THAIGEKNQLEVSVTLSPOBWSLSQISDTGKKNRFTVTGGRGGSGHPMVTVASDIAR 178
Db 117 KHIAVGQNGVETSVLSVDQYARLQSIDPEGDKFVFTGGRGGAGHAMVTVASDITEAR 176
Qy 179 TRILAKLPDNGHGRQPKVDVTRSVGVCSASGI-----DDGV--VSTHTSTNVSYS 230
Db 177 QRILLELEPKGTG-----ESKGAGESKGVGELRESNGAENTTTQTSTSTSLRS 227
Qy 231 DPKEVWSVGAIAAGLAGLAATGIAQALALTPEPDDPTTTDPQAAANAESATKQDLTOBA 290
Db 228 DPKLWALGTVATGLIGLAATGIVQALALTPEPSPPTTTDDPAAASATETATRDULTKEA 287
Qy 291 FKNPENQKVIDANGNAIPSGELXDDIVEYIAQQAQKEAGEVARQQAQVSNQAQQRVEDQ 350
Db 288 FQNPQNKVNDLGNALPSGVLDVVANIEEQAAGEAKQAIENNAQAQKQYDBQ 347
Qy 351 HARQEEQLSSGGYGLSSALIYAGGIVAGVTALHRRNQPAEQTTTTT-----HT 403
Db 348 QAKRQEELKSSGAGYGLSGALILGGGIGVAVTAAALHRRKNQFVEQTTTTTTTTTSART 407

Qy 404 VVQQTGGIPQHVKALMPQERRRFSRRDSQGSVASTHWSOSS-SEVVNPYAEVGGARNS 462
Db 408 VENKPNANTPAQGNVDTPGSEDTMESRRSSMASTSTFTFDTSIGTVQNPYADV-----K 462
Qy 463 LSAHQPEEHIYDEVAADPGYSVIQNFGSGGPVT-----GRLLGTPGQGIQ 507
Db 463 TSLHD-----SQVPTSNTSVQMGNTDSVVYSTIOHPPRDPTDNGARLLGNPSAGIQ 516
Qy 508 STYALLANGSGLRLGCGGLTSGGTAVSSVNAAPTGPVRFV 549
Db 517 STYARLALSGGLRHDHMGGLTGGNSAVNTSNPPAPGSHRFV 558

RESULT 2
E86045
probable translocated intimin receptor protein tir [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E86045
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: E86045
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-558 <STO>
A:Cross-references: GB:AE005174; NID:g12518449; PIDN:AG58825.1; GSPDB:GN00145; UWGP:Z51
A:Experimental source: strain O157:H7, substrain EDI933
C:Genetics:
A:Gene: tir

Query Match 54.6%; Score 1528.5; DB 2; Length 558;
Best Local Similarity 56.2%; Pred. No. 6.5e-88;
Matches 327; Conservative 60; Mismatches 138; Indels 57; Gaps 11;

Qy 1 MPIGNLGNVNGNHLIPAPPPLPSQTDGAARCGTGHLISSTGALGSRSLFSPLRNSMADS 60
Db 1 MPIGNLGNPNVNSIPAPPPLPSQTDGA--GGRQLINSTGPLGSRALFPPVRNSMADS 58
Qy 61 VDNR--DIPGLTPNRSRLAAATSETCLLGGFVLHDKGFLDILNTQIGPSAFRVEQADG 118
Db 59 GDNRASDVPLGPNVPMRLAA--SEITLNDGFEVLHDKGFLDILNRQIGSSVFRVETQBDG 116
Qy 119 THAAIGKNGLEVSVTLSPQWSSLOSIDTEGKNRFVFTGGRGGSHPMVTVASDIAEAR 178
Db 117 KHIAVGQRNGVTSVVLSDQEVARLQSIDPEGKDFVFTGGRGGAGHAMVTVASDITEAR 176
Qy 179 TRILAKLPDNDHGRQPKDVTDRSVGVGSASGI-----DDGV--VSETHSTTNSSVRS 230
Db 177 QRILELLEPKGTG-----ESKGAGESKGVGELRESNSCAENTTETQTSTSTSLRS 227
Qy 231 DPKFWVSGAIAAGLAGLAATGIAQALATPDPDPTTDPDQAANAESATKQOLTOEA 290
Db 228 DPKMLALGTVAATGLIGLAATGIVQALATPDPDPTTDPDQAASAETATRDQOLTKEA 287
Qy 291 FKNPENOKNIDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQOAVESNAQAQRYEDO 350
Db 288 FQNPONOKNIDELGNAIPSGVLKDDIVANIEEQAKAAGEEAQQAENNAQAQKXYDEQ 347
Qy 351 HARRQELQSSGIGYGLSSALIVAGGIGAVGTALHRRNQPAEQTTTTT-----HT 403
Db 348 QAKRQELKSSGAGYGLSGALLGGIGVAVTAALHRRKNQVQETTTTTTTTTTSART 407
Qy 404 VVQQTGGIPQHVKALMPQERRRFSRRDSQGSVASTHWSOSS-SEVVNPYAEVGGARNS 462
Db 408 VENKPNANTPAQGNVDTPGSEDTMESRRSSMASTSTFTFDTSIGTVQNPYADV-----K 462
Qy 463 LSAHQPEEHIYDEVAADPGYSVIQNFGSGGPVT-----GRLLGTPGQGIQ 507
Db 463 TSLHD-----SQVPTSNTSVQMGNTDSVVYSTIOHPPRDPTDNGARLLGNPSAGIQ 516

Qy 508 STYALLANGSGLRLGCGGLTSGGTAVSSVNAAPTGPVRFV 549
Db 517 STYARLALSGGLRHDHMGGLTGGNSAVNTSNPPAPGSHRFV 558

RESULT 3
B90784
hypothetical protein ECa1242 [imported] - Escherichia coli (strain O157:H7, substrain RI
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B90784
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
sasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: B90784
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2793 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA34665.1; PID:g13360702; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 050952
C:Genetics:
A:Gene: ECa1242

Query Match 5.4%; Score 151.5; DB 2; Length 2793;
Best Local Similarity 21.4%; Pred. No. 0.61;
Matches 115; Conservative 87; Mismatches 222; Indels 113; Gaps 27;

Qy 22 LPSOTDGAARG--GTGHLISSTGALGSRSLFSPLRNSMADSVDSDRDLPLGTPNPSRLAAA 79
Db 55 LPTMGKGFQSVRGTEGEM--ARGLGDAITQSPVKTG--ARILNEFSRMGLP-----GVA 104
Qy 80 TSETCLLGGFVLHDKGFLDILNT-QIGPSAFRVEV---QADGTHAAIGKNGLE-VSV 133
Db 105 TVQIDIFAGG-----SRGADEVIDTLPDGKRAVTTVGKLGKATGKAVSDGAKATDEWLTG 159
Qy 134 TLSPQWSSLOSIDTEGKNRFVFTGGRG---GSGHP-MVT--VASDIAEARTR-----180
Db 160 KMSFGAVRALNTPMTEGVNDSAVVAVKGNVLIGALVPMVAGGVARKVGDVTLRKLMTAG 219
Qy 181 -----ILAKLPDNDHGRQPKDVTDRSVGVGSASGIDGVSSETHSTTNSSVRSRDPKFW 235
Db 220 LEKYIYAAGMQPERATALAAEAVDKMPDLFQA-----GLITHSTVAQGS-----266
Qy 236 VSVGAIAAGLAGLAA--TGIAQA-----LALTPEDDPTTDPDQAANAESATKQOL 286
Db 267 ----AMAAADAVLNADYSELQAQSPKFOQTFLSIDADPQHAQLTD-RQWDLAKERVADEV 321
Qy 287 TOEAFQKPNOKNIDANGNAIPSGELXDDIVEQIAQQAKE--AGEVARQOAVESNAQAQ 344
Db 322 RAQLATDPELLAVN--AMAAKLGDAQLFNLVTRGTAKTVKSGIVRNATEQGAINAAQGGY 379
Qy 345 QRYEDQHARRQEEELQSSGIGY----GLSSALI VAGGIGA-----GVTALHRRNQPAE 394
Db 380 SRYQENTALRE-----TAGMGVSPWEGVADATIEGAAFGAAMGAPFGGAVAGYGRRRQAAE 434
Qy 395 OTTTTTTTVVQQTGGIPIQ-HKVALMPOER-----RRFSDR-RDSQGSVAST 440
Db 435 ETAMRDAETV--QODDRAAPQESVDPAVQAQRESQGMNRQQLLEQYADADNATEGDASAA 492
Qy 441 HWSDSSESVNPNYAEVGGARNSLSAHQPEEHIYDEVAADPGYSVIQNFGSGGPVTGR 497
Db 493 HRREAASQLL-----NELDQTKRAQVMNELKAKPRSELLEEYRRLSQKEGR 539

RESULT 4
A28182
hemolysin A - Serratia marcescens
C:Species: Serratia marcescens
C:Date: 30-Jun-1989 #sequence_revision 30-Jun-1989 #text_change 08-Oct-1999
C:Accession: A28182
R:Poole, K.; Schiebel, E.; Braun, V.

J. Bacteriol. 170, 3177-3188, 1988
A:Title: Molecular characterization of the hemolysin determinant of *Serratia marcescens*.
A:Reference number: A28182; MUID:88257037; PMID:3290200
A:Accession: A28182
A:Molecule type: DNA
A:Residues: 1-1608 <POO>
A:CROSS-references: GB:M22618; NID:g340726; PIDN:AAA50323.1; PID:g556420

Query Match 5.4%; Score 150.5; DB 2; Length 1608;
Best Local Similarity 20.9%; Pred. No. 0.32;
Matches 122; Conservative 84; Mismatches 236; Indels 141; Gaps 25;

QY 33 GTGHLISSTGALGSRSLFSPRLNSWADSVSDRDIPLGTPNPSR-----LAAATSETC 84
DB 1008 GSDLTVDAKGEQGTQSRSSASQAQVTSIDAA--GINVVKDAIYQQTALNGRGKTA 1065
QY 85 LLGGFEVLHDKGPLDILNTQIGPSAFRVEVQA-----DGTG----- 120
DB 1066 VNAGDRIQLDQASDKQSESR-----SGFNKASAKGFTADSKNFGAGFGGTHNGSSSST 1122
QY 121 AAIGEKNGLEVSVTLSPOEWSLSQSIDTEGKRNRFVFTGGRGSGHGMVTVASDIAEA-RT 179
DB 1123 AQVGNISG-QQGVELKAGRDLTLOQTDVKSQGDVLSAGN-----KVALQAESTQT 1173
QY 180 RILAKLDPNHHGGRPKVDTRSVGVGSAG-IDGVVSEHTSTNSSVRSDPKFWVSV 238
DB 1174 RKESKLSGNIIDLGAGSSDSKENTGNLSAGGAFDIAKVNESATERQGIATIASDGKVTLSA 1233
QY 239 ---GAIAAGLAGLAATGTAQALALTPEDDPTTDPDQAANA---ESATKDOLTQEAFF 292
DB 1234 NKGDDALHLOQAKVSGGSAAL-----EAKNGGILLESAKNEQ----- 1271
QY 293 NPEN-----QKNVIDANGNAIP-----SGELKDDIVEQIAQAAK-----E 327
DB 1272 HKDNWSLGTKANAKGGQFNKAGKVDNKGKDTHTLGLAGLVKGVQDDKTHANTGIT 1331
QY 328 AGEVARQAQAVESNAQOQRYEDQHARRQBELQLSSGIGYGLSSALIVAGGIGAGVTTALH 387
DB 1332 AGDVTLSNGKDTRL-AGARVD-----ADSVQGVKGGDLHVSRRKDVGVKVDVDAGLS 1384
QY 388 RNQPAEQTTTTTTHVVQOQTGPIQ-----HKVALMPQER-----RRFSDRDSDGS 436
DB 1385 HSNDFGSSITSKSLKVTGPYAGKVKLEAGNVKADATTDKYNVARRLDPQODTTGA 1444
QY 437 VASTHWSSSSVNPNYAEVGGARNLSAHOPEHHIYDEVAADPGSVIQNFGSGPVTG 496
DB 1445 VS---FSKAEKVTLPATPAG-----EKGQGLWDRGARTVGVGAVKDSITGPAGRQG 1493
QY 497 RL-----IGTPGQGIQSTYALLANSGLRLGMGGLT--SGGE 531
DB 1494 HLKVNADVNNNAVGEQSA---IAGKNGVALQVGSGTQLTGGE 1533

RESULT 5
C71513
hypothetical protein CT456 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: C71513
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
A:Reference number: A71570; MUID:9900809; PMID:9784136
A:Accession: C71513
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1005 <ARN>
A:CROSS-references: GB:AE001319; GB:AE001273; NID:g3328881; PIDN:AAC68056.1; PID:g332888
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: CT456

Query Match 5.4%; Score 150; DB 2; Length 1005;

Best Local Similarity 22.8%; Pred. No. 0.18;
Matches 145; Conservative 65; Mismatches 234; Indels 192; Gaps 33;
QY 4 GNLGNVNGNHLIPAPPL-PSQTDGAARGT-----GHLISSTGALGSRSLFSPRL-- 54
DB 431 GNLNTVLIINNKFKTCVAYGPNWQSEASSGYTPSAWRGRHVRDFGGEIKANDFNKINWG 490
QY 55 -----NSWADSV--SRDIPGLFTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIGPS 108
DB 491 TQAGPSSDDGIGSPNETPG--AGP---AAAPSPT-----PSSIIPII----- 528
QY 109 AFRVEVQADGTHAIGEKNGLEVSVT--LSPOEWSLSQSIDTEGKRNRFVFTGGRGSGHPM 167
DB 529 ---VNVVGGTNNVIGDTN---VNTNTTPTTQSTDASTD- 563
QY 168 VTVAADIAEAATRLAKLDPNHHGGRPKVDTR---SVGVG-----SASGIDDG 214
DB 564 ---SDIDDIINT-----NNQTDINTTKDSGDGAGVNGDISETESSSGDDSG 607
QY 215 VVSEHTSTNNSVRSDPKFWVSV-----GAIAAG-----LAG 247
DB 608 SVSSSE--SDKNASVGNDCGPAMKDILSAVRKHLDVVYPCGNGSGTEGPLPANQTLGDVISD 666
QY 248 LAATGIAQALALTP-----PDDPTTDPDQAANAESAT-----KQULTQEA----- 290
DB 667 VENKGSADOTKLSGNTGAGDDDDPTTT--AAVGNGAEEITLSDTDSGIGDDVSDTASSGD 724
QY 291 ---PKPENOKVINANGNAIPSGELKDDIVEQIAQAAKEAGEVARQAQAVESNAQAQOR 346
DB 725 ESGGVSSPSSSENKNTAVGNDGPSGL---DILAAVRKHLDKVYPCDNGSGSTEGPLQANQT 781
QY 347 YED-----QHARROEEQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTT 401
DB 782 LGDIVQDMETTGTSQETVVSF--WKGSTSTSESAGSGSVOT--LLSPPTPTTTLRT 837
QY 402 HT-----VVQOQTGG-IPQHKV--ALMPQERRRPSDRRDSQGSVASTHW 442
DB 838 CTGATTTTSLAMGGPIKADIITGGGRIPGGTLEKLLPRIRAHLDISFDAGQGLVSTEE 897
QY 443 SDSSEVVNPNYAEVGGARNLSAHOPEHHIYDEVAADPGSVIQNFGSGPVTGRLIGTP 502
DB 898 POLGS-IVNKFRQETGSRGILA-----FVESAPGRPESA--QVLTGTG-----GDK 940
QY 503 QGGIQTSTYALLANSGLRLGMGGLTSGGETAVSSVN 538
DB 941 GNLQAAAAVTQALGNV-AGKVNLAIQGKLSLNV 975

RESULT 6
DB5644
hypothetical protein Z1495 [imported] - Escherichia coli (strain O157:H7, substrain EDL9
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: DB5644
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Griest, E.J.; Davis, A.; Lim, A.; Dimalanta, E.; Potamianos, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: DB5644
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-2806 <SPO>
A:CROSS-references: GB:AE005174; NID:gl2514354; PIDN:AAG55616.1; GSPDB:GN00145; UWGP:Z14
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1495

Query Match 5.3%; Score 149.5; DB 2; Length 2806;
Best Local Similarity 20.8%; Pred. No. 0.82; Mismatches 218; Indels 123; Gaps 27;
Matches 113; Conservative 88; Mismatches 218; Indels 123; Gaps 27;
QY 22 LPSQTDGAARG--GTGHLISSTGALGSRSLFSPRLNSMADSVSDRDIPLGTPNPSRLAAA 79

Db 68 LPTMGKGAQSVRGTEG---ARGLGDAIQSPVKTG-ARILNEFRMGLP-----GVA 117
Qy 80 TSETCLGGEFVLDKGLDILNTQ-----IGPSAFRVEVQADGTHAAIGKNGL 129
Db 118 TVQDIFAGG-----SRGAEVIDTLDPGKNAVTDIVGKG-----LKATGKAVSDGAKATD 167
Qy 130 E-VSVTLSPQEWSSLSQSIDTEGKRFVFTGGRG-----GSGHP-MVT--VASDIAEARTR- 180
Db 168 EWLTKMSPCAVRALNTPMTEGYNDASVWVAKGNLIGALVDMVAGGVARKVGDVTLRK 227
Qy 181 -----ILAKLPDNDHGRQPKVDTRSVGVSGSAGIDGCVVSETHSTTTNNSVRS 230
Db 228 MLTAGLEKKYIAAGHOPERATALAAEAVDKMPDLFOA-----GLITHSTVSAOQS--- 279
Qy 231 DPKFWVSVCAIAGLAGLAA--TGTAQA-----LALTPEDDPTTTPDQAANAESA 281
Db 280 -----AMAAADAVLNADYSELAQSPKFOOTLSLADADPOHAQLTD-RQMDLAKER 329
Qy 282 TKDQLTOEAFKNPENKKNIDANGNAIPSGELKDDIVEQIAQOAKE--AGEVARQOAVES 339
Db 330 VADEVRAQLATDELLAVN--AMAAKLGDAQLFNLVTRGTAKTVKSGIVRNATEQAINA 387
Qy 340 NAAQOQRYEDQARQEBELQSSGIGY-----GLSSALIVAGGIGA-----GVTTALHRR 389
Db 388 AQQGYSRYQENTALRE-----TAGMGVSPWEGVADATIEGAAGFAMGAPFGAVAGYRGR 442
Qy 390 NQPAEQTTTTTTHVVQQTGGIQQ-HKVALMPQR-----RRFSDR-RDSOG 435
Db 443 ROAAEETAMRDAETV--QDDDAAPQESVDPAQOQRESQGMNRRLQLEQYADADAMATEG 500
Qy 436 SVASTHWSDSSEVNPVYAEVGVGARNLSAHQPEEHYDEVAADPGYSVIONFSGSPVT 495
Db 501 DASAHRREAASQLL-----NELDEQTKQAVNMELKAKPRSELLEBYRLSQKE 550
Qy 496 GR 497
Db 551 GR 552

RESULT 7
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34434
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: EMBL:U80846; PIDN:AAAC70890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 2075/1

Query Match 5.3%; Score 149; DB 2; Length 2232;
Best Local Similarity 19.9%; Pred. No. 0.64;
Matches 112; Conservative 71; Mismatches 238; Indels 142; Gaps 19;

Qy 21 PLPSQ--TDGAARGGTHLISSTGALGSRSLFPLNSMADSVDSRIDPG-----LPTNPS 74
Db 899 PPSQSTTIGSTQGSTSPGISTT-----SEEMTSQGSTQTPGSGSTVTQPS 945
Qy 75 RLAAAT--ETCLLGFEVLHDKGPLDILNT-----QIGPSAFRVEVQAD 117
Db 946 TVSDSTSSGSGTVTVGTEGSSPIPTSONTPNPSSTSGSSMSTQTPQSSQSTSPVESSTS 1005

Qy 118 GTTHAIGCKNGLEVSVTLSQPEWSSLSQSIDTEGKRFVFTGGRGSGHPMVTVASDIAEA 177
Db 1006 GATSSSGSPGTLTISPSPPSSTI-----GSSQGSTPVV---STISQ 1048
Qy 178 RTRLAKLPDNDHGRQPKVDTRSVGVSGSAGIDGCVVSETHSTTTNNSVRSDDPKFWVS 237
Db 1049 STE-----TPGSGSTVTKP-----STVSGSAGSGSTATMGSTASSTSGSGSSTPNPSQS 1099
Qy 238 VGATAAGLAGLAATGIAQALALATPEDDPTTTPDQAANA--AESATKQQLTOEAFKNPE 295
Db 1100 TSPSTSGATSPSGSGTLTISPSPSQSTIGSSQGSTFPVVTSTGDMTSQGSTQIPG 1159
Qy 296 NQKYN-----DANGNAIPSGELKDDIVEQIAQO--KEAGEVARQOAVSNAAQOQRYE 348
Db 1160 STGSTVTPQSTGSGSTSGEITTSQSTQTPRSLSTSPAISTSQGSVSTNSP-----1213
Qy 349 DOHARRQEBELQSSGIGYGLSSALIVAGGICAGVTTALHRRNOPAEOTTTTTTTHVVQOQ 408
Db 1214 -----GSTVTPQSTVGRSTSGSTVITGSGSST 1243
Qy 409 TGGIPQHKVALMPQERRRFSRRDSQGSVASTHWSDSSEVNPVYAEVGVGARNLSAHQP 468
Db 1244 SGS-----SSATSLSSSPVPSTSQSPN-----STSGSSTPTPNP 1279
Qy 469 EEHIYDEVAADPGVSVIONFSGSGPVTGRLIGTPGQGIQSTYALLANSGLRLMG--G 525
Db 1280 SQSTSPVVTGSGTSGTSGTSTGVTQ-----STVS--GSNSSGSGSTVTIGSSEA 1333
Qy 526 LTSGG--ETAVSSVNAAPTQCPV 546
Db 1334 STSGSPKTPSSISPVTSSPI 1356

RESULT 8
T24583
hypothetical protein T06D8.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T24583
R:Palmer, S.
submitted to the EMBL Data Library, April 1995
A:Reference number: Z1909
A:Accession: T24583
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1829 <WIL>
A:Cross-references: EMBL:Z49130; PIDN:CAA88964.1; GSPDB:GN00020; CESP:T06D8.1
A:Experimental source: clone T06D8
C:Genetics:
A:Gene: CESP:T06D8.1
A:Map position: 2
A:Introns: 1391/3; 1432/3; 1470/3; 1505/1; 1520/1; 1616/1; 1644/1; 1687/3; 1742/1

Query Match 5.2%; Score 146; DB 2; Length 1829;
Best Local Similarity 21.7%; Pred. No. 0.74;
Matches 118; Conservative 63; Mismatches 256; Indels 108; Gaps 19;

Qy 68 GLPTNPSRLAAATSETCLLGFEVLHDKGPLDILNTQIGPSAFRVEVQADGTHAAIGKN 127
Db 172 GEETTTSVATEASSEA-----TTTPAGTEASGEET---TTSVATGS 210
Qy 128 GLEVSV-----TLSPQEWSSLSQSIDTEGKRFVFTGGRGSGHPMVTVASDIAEARTRI 181
Db 211 GEETTVAVVESSGEBEPASSTSIPTLSKNDQV-----TEASGETITATAATEASEETTT 266
Qy 182 LAKLPDNDHGRQPKVDTRSVGVGSAG-----IDGCVVSETHSTTTNNS 227
Db 267 SAVTEGSGE-----DTTVAVVVELSGEQPASSTSIPTLSKDDQVTEASGETTTA 318
Qy 228 VRSDPKFWVSVCAIAGLAGLAATGIAQALALATPEDDPTTTPDQAANAESA 281
Db 319 ANTEASETTTTSVATEG--SGEETTVAVVESSGEBEPASSTSIPTLSKDDQVTEASGEE 377

[illegible]

RESULT 11
T42233
submaxillary mucin 1 - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Nov-2000
C:Accession: T42233
R:Jiang, W.; Woitach, J.T.; Keil, R.L.; Bhavanandan, V.P.
Biochem. J. 331, 193-199, 1998
A:Title: Bovine submaxillary mucin contains multiple domains and tandemly repeated, non-
A:Reference number: 222092; MUID:98180948; PMID:9512473
A:Accession: T42233
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-1589 <JIA>
A:Cross-references: EMBL:AF016589; NID:g3057086; PID:g3057087; PIDN:AAC39250.1
C:Genetics:
A:Gene: SM1
C:Superfamily: big submaxillary mucin

Query Match	5.1%	Score 144;	DB 2;	Length 1589;
Best Local Similarity	19.5%	Pred. No. 0.81;		
Matches 119; Conservative	87;	Mismatches 247;	Indels 156;	Gaps 25;
QY	7	GNVYNGNHLIAPPPLPSQTGGAA----	RGCTGHLISS-----	TGALGSRSLFSPLRN 55
Db	66	GTNPV-----VSGAPVTPGSSAGSGCAPOTGGPGSETASPLSGAAGTATGSGRTSIPP----	118	
QY	56	SMADSVDSRDIPGLPTNPSPRLAAATSETCL-----	LGCFEVLHDKGPDILNLTQIGPS----	108
Db	119	--SGAPVTPPEPLISTGASAGPPASSESTVTLPGATGTDVLRSCTSLPVSQCAVTPPASS	176	
QY	109	-----AFRVEVQA-----	DGTHAAIGKNGLEVSVTLSPOEWSLQ	144
Db	177	CGSSATAGPAVGSQTTVQVSGAFATHVAKSNNTNGSSAEIS	TTTGATAGLTLTSE--TSIV	234
QY	145	SIDTEGKNRFV-----	FTGGRGSGHPMVTVASDIAEARTILAKLPDND--HG	191
Db	235	SSATRAPSSAVTRAPVTVDDVSGISHSSGSRRTIVIGPSSVSSAGAEIAPS	SLGTDGLEG	294

Qy	192	GROPKVDTRSV---- <div>GVCSAGIDGVVSETHSTTNSVSRDPKFWVSVGAIAAGLAG</div> <div>247</div>
Db	295	TTKISVDARTIRPSYGALGATGSSIGEIGTTST-----PEF-TETTSFSVGLR- <div>343</div>
Qy	248	LAATGIAQALALNPEDD-PTTTDPQANAAESATKDQLTOEAFKNPENOKNVINDANGN <div>306</div>
Db	344	-----TTRSSSETGTTLESSTASSSSSTGTTGSIAGLRNTRISLRSRSGT <div>391</div>
Qy	307	AIP-SGELXDDIVEQIAQAQAKEAG-----EVARQQAQVESNAQAQRYVEDQHARR <div>354</div>
Db	392	TRPSSGETETTVIESRVSGSDEGLCTIGSTAGLMTTRTRISVVVSGTGTSPSSSETGSA-- <div>449</div>
Qy	355	QEBLQISSGIGYGLSALIVAGGIGACVITALLHRRNQPAEQTTTTTHTVVQOQTGGIPQ <div>414</div>
Db	450	VSEFRTSGSLGKG-SRTTWTSTPOL-ARMTRISFGSGTTRQSSGETGTTTVIESRTSGSP- <div>506</div>
Qy	415	HKVALMPQERRRDRDSSGVSASTHWSDSSEVNPVPAEVCAGARNLSAHOPEEHYD <div>474</div>
Db	507	-----SEGLRTGSTAGL-----TRTII-----SVGSAATEPSSRTEYTVTE <div>546</div>
Qy	475	EVAADPCYSVIQNFSGSGPVTRGLITPGQGIQSTYALLANSGLRLGMGG-----LTSGG <div>530</div>
Db	547	-----SNNGSLGEGSGTTCGAIAGLTTTRTRISVVSGTTRPSSG <div>585</div>
Qy	531	ETAVSSVNA 539
Db	586	ETRTTVIES 594
RESULT 12		
B98047		
hypothetical protein spr1403 [imported] - Streptococcus pneumoniae (strain R6)		
C;Species: Streptococcus pneumoniae		
C;Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001		
C;Accession: B98047		
E;Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoffe, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhrr, Y. P.; Sun, P.M.; Winkler, M.E.		
J. Bacteriol. 183, 5709-5717, 2001		
A;Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskun, A.;Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.		
A;Reference number: A97872; MUID: 21429245; PMID:11544234		
A;Accession: B98047		
A;Status: preliminary		
A;Molecule type: DNA		
A;Residues: 1-2551 -KUR-		
A;Cross-references: GB:AE007317; PIDN:AAL00207.1; PID:g15459054; GSPDB:GN00174		
C;Genetics:		
A;Gene: spr1403		
Query Match		
Best Local Similarity 5.1%; Score 143.5; DB 2; Length 2551;		
Matches 131; Conservative 75; Mismatches 245; Indels 165; Gaps 32;		
Qy	7	GNVNGNHLIPPAPLPSPQTDGAARGGTGHLISS---TGALGS-----RSLFSPPLNSM 57
Db	1724	GKGNFSHTIVTF-YTDENNNGKYVTPCTDELLGSEMIKDCAGADGRDGSLLT-VKDGK 1781
Qy	58	ADSVDSRD--IPGLPTNPSP-----LAAATSETCLLGGFEVL---HDKGP 97
Db	1782	ETKYVQEDPANPQPLNPEKPLAVIRDVGDKSPVTAVRKBAGHKGVEITVDNHDGSG 1841
Qy	98	LDILNTQIGSAPRVEQADGTHAATGEKNGLSVTLSPQE-WSSLSQISDTGKRFVF 156
Db	1842	-----PTTVFVQDGAKGKTGATGQDQGTPTITTRQGDGGSTVVTITTSKQPVTF 1892
Qy	157	T---GGRGGSGH-PMVTVASDIAEARTR---ILAKLDPDNHGRQPKDQDTRSVGV----- 205
Db	1893	TVKDGKNGKDGRAFKIKVEDITSPSRIRRDTHAATPRNGIRVTVYDDYDNDNGVYDEGV 1952
Qy	206	-----GSASGI---DDGVVSETHSTTNSVSRDPKFWVSVGAIAAGL 245
Db	1953	DKVLNSKDIYNGIDGRGSAPTITTKONG--DGTHITTVQNPQDGE-----STTVVKDGK 2005

Qy 144 QSIDTEGKRVFTTGGRGSGHPMV-----TVASDIAEPARTRIIAKLPDPNHHGRQP 195
Db 152 EEISNK-----PPVDEESESASDCREFKLVNLHQOQH--HSP 194
Qy 196 KDVP-THSVGVGSAGIDD--GVYSEHTSTTNSSVSFDPKFVWSVAIAGAAGLAATG 252
Db 195 SSPDKTRST-----LDDVKILWERKKQLQRSSV----- 223
Qy 253 IAQALALTPE-----PDPPTTTDPDAANAASATKDOLTQEAFKNPKNOKVNIIDA 303
Db 224 ITAAPTLPQOHQPMDSIEDIEETLEDVDADAVEADAEELEEQYNGYDSPDLDSL 283
Qy 304 NG-----NAIP-----SGLXDDIV--EQIAQAQKEAGEVARQQAVE----- 338
Db 284 GGATSAASAASAAVASRRRGRTRYSGTESDQAACERARMRLKPERKAERSAAYKKSLM 343
Qy 339 -----SNQAQORYEDQHAROEELQLSSGIGYLSSALI VAGGIGAGVT 384
Db 344 KRYYTEIPVKQSTSPAQOOLQOQHLLQOQQOQPHGSTF-----AGATA 390
Qy 385 ALHRNQPAEQTTTTTHVVQOQTGGIPQHKVALMPQERRRFSDRRDSQGSVAETHWSD 444
Db 391 LLHIK-----TEQNTLLPLQLQOQO-----QQOQG----- 416
Qy 445 SISEVNYPAYBVGARNLSAHQPEEHYDEVAADPGYSVIQNFSGGPV----- 494
Db 417 -----LHCAGNGSSGNNAHQOQPL-----AIQRPPLLHLNLSSGAHNPHRNYYT 466
Qy 495 --TGRLIGTPG-QGIQSTYALLANGS-----LRIGMGLTSGGETAVSSVNA 540
Db 467 ATTGFPSPADSGVDVDS--SSSGQPCADELKARLGMPPATASAAAAA 521

RESULT 11
T42233
submaxillary mucin 1 - bovine (fragment)
C:Species: Bos primigenius taurus (cattle)
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Nov-2000
C:Accession: T42233
R:Jiang, W.; Woitsch, J.T.; Keil, R.L.; Bhavanandan, V.P.
Biochem. J. 331, 193-199, 1998
A>Title: Bovine submaxillary mucin contains multiple domains and tandemly repeated, non-
A:Reference number: Z22092; MUID:98180948; PMID:9512479
A:Accession: T42233
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1589 <JIA>
A:Cross-references: EMBL:AF016589; NID:g3057086; PID:g3057087; PIDN:AAC39250.1
C:Genetics:
C:Gene: SM1
C:Superfamily: pig submaxillary mucin

Query Match 5.1%; Score 144; DB 2; Length 1589;
Best Local Similarity 19.5%; Pred. No. 0.81;
Matches 119; Conservative 87; Mismatches 247; Indels 156; Gaps 25;

Qy 7 GNNVNGNHLPAPPLPSQDTGAA-----RGGTGHLISS-----TGALGRSLFSPURN 55
Db 66 GTNVP-----VSGAPVTPGSSAGSSGAPGTGPGSETASPLSCAAAGTSATGSRITIPP--- 118
Qy 56 SMAUSVDSRDIPGLPTNSRLAATSETCL-----LGGFEVLHDKGPLDILNTQIGPS--- 108
Db 119 --SGAPVTPPEPLISTGASAGPPASSSETVTLPGATGTDVILRSGTSLPVSGCAVTPASS 176
Qy 109 -----AFRVEVOA-----DGTHAAIGEKNGLEVSVTLSPOEWSLQ 144
Db 177 CGSSNATGAPVGSQTTVQVSGAFATHVKAINTNGSSAEISSETGATGLTITSE--TSIV 234
Qy 145 SIDTEGKRVF-----FTGGRGGSGHPMTVVASDIABARTIRILAKLPDN-HG 191
Db 235 SSATRAPPASVTRAPVTYDDVSGISHSSGHSRTIVIGSPSSVSAEQIAPSLSLTDGLEG 294

Qy	375	AGGIGAVTTALHRRNQ----	PAB-----QTTHVTVVQQQTGGIPQHVKVALMPQER	421
Dd	437	VGG-----TAAEHVPQAVTAFTETAHD	FSEKTTTASKRVSKDASA8KHVKARPK---	488
Qy	425	RPFSDRDSQGSVASTHWSSSSVNPNPYAE	VGVGARNLSAHQPDEHIYDEVAADPGYSV	484
Dd	489	-----SSTGQEPPTTPAKSAQS	KHARRSKQ-----ASNAP-----	521
Qy	485	IQNPSGSGPVTGRLLTGPGGIQSTYALLANS	GGLRLGMGLTSGGETAVSSVNAA---P	541
Dd	522	----SSPGTTTAAVPGCKKSAIEAA-APIPTS	ADTVESKHAAGSGSATTFSPGSAATTKP	576
Qy	542	TPG 544		
Dd	577	TPG 579		
RESULT 14				
AC0304				
Probable hemolysin YPO2490 [imported] - Yersinia pestis (strain CO92)				
C;Species: Yersinia pestis				
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001				
C;Accession: AC0304				
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, geno-Tarraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dou-				
il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.;				
Nature 413, 523-527, 2001				
A;Title: Genome sequence of Yersinia pestis, the causative agent of plague.				
A;Reference number: AB0001; MUID:21470413; PMID:11586360				
A;Accession: AC0304				
A;Status: preliminary				
A;Molecule type: DNA				
A;Residues: 1-2535 <KUR>				
A;Cross-references: GB:AL590842; PIDN:CAC91295.1; PID:g15980484; GSPDB:GN00175				
C;Genetics:				
A;Gene: YPO2490				
Query Match 5.0%; Score 141; DB 2; Length 2535;				
Best Local Similarity 21.9%; Pred. No. 2.4;				
Matches 129; Conservative 73; Mismatches 246; Indels 140; Gaps 25				
Qy	25	QTDCA-----ARGGTGHLISSTGALGSRSLF	SPLRNSMADSVDSDIPGLPTNPRLAAATS	81
Dd	1485	QTDGALTLAGGDVLLTTQTEQDHEQRNHTGL	SKGIASLTTRTDSL-----S	153
Qy	82	ETCLLGFEVLHKPDILNTIGCPAFRVEQAD--	-----GT-----HAA	122
Dd	1534	QTLAVGS---MLSAGSIDVSGKNIAVMGSNV	VADQDISLRAQENITVGTAQOSESHEL	1599
Qy	123	IGEXNL-----EVSATL---SPOBSSLOSID	TGEKNRFVTGGGGS--GHPMVTVASD	173
Dd	1591	EQKKSGLMSTGIGIVTGVSSSTKMVDSQSISS	-----VGSTVGSVLGNVSMTAGED	164
Qy	174	IAEARTRILAKLDPPNHGGR-----	--QPKDVDTFRSVGVGSA-SGIDDGVVSETH	221
Dd	1643	LRVQGAELVLAGKDINLTGKNVSIILAEENOL	TQSHVEQKGSLTLALSAGVSAVNTAVT	170
Qy	222	STTNSSVRSDPKFWVSWGAI AA--GLAGLAT	GIAQAALALPEDDPTTDPDQANAAE	279
Dd	1703	TAKAASESSRGICALQVKKAALNGVQAVQAGL	VQA-----EGDDAASMFGISASLGSQ	175
Qy	280	SATXDO-----LTQEAFKPENQKVNDANGNAI	PSGELXDDIVEQIAQAQAKEAGEV---	331
Dd	1758	KSSSEQHQEQHTVGTSTLTAGNNITINATGEGNA	NSG-----DIVVQ-GSQLAQAGDRTL	181
Qy	332	-ARQAVESNAQAQRYEDOHARQEBELQLSSGI	GYGLSSALI VAGGICAGVTTALHRRN	390
Dd	1813	DAARDVLLGNANTQKTGSSNSGGSVCSVLGIS-	GASSGLSIFANANKG-----QGS	1861
Qy	391	QPAEQTTTTTHTVVQQQTGGIPQHVKVALMPQ	ERRRFSDRDSQGSVASTHWSDS---S	447
Dd	1866	EHGDTGTSWTET-----	-----TLDSSGTLISLYSGRDTSLVCA	189

Search completed: December 16, 2003, 08:56:49
Job time : 30.2608 secs

```
QY 448 EVNPNYAEVGARNLSAHQPEEHYDEVAADPGVSVIQNFSGSGPVTGRLLIGTPGQ--- 504
Db 1898 QVSGETVKEVGRDILLQSDSDNYD--AKQNSSVGSGSFS-PCSMGTGSIINGSQDKL 1954
QY 505 -----GIQSTYALLANSGLRLGGLTS-----GGETAVSSVNAAPT 542
Db 1955 HSNFDSVQEQTGIFAGSGGFDITVGGHTQLDGAIVIGSTATADKNTLDT 2002

RESULT 15
AE1852
Hypothetical protein alr0366 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AE1852
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AE1852
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1152 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA072324.1; PID:gl129711; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr0366

Query Match 5.0%; Score 140; DB 2; Length 1152;
Best Local Similarity 21.2%; Pred. No. 0.91;
Matches 127; Conservative 90; Mismatches 236; Indels 146; Gaps 29;

QY 5 NLGNVNGNHLPPAPPILPSQTDGAARGTGHLISSTGALGSR-----LPSPLNSMADS 60
Db 124 NASLNIIGSFVATTANTI--QPPGAEFSLTSPVTSTNTLLSVNPTAFLEQNIAQNGTNS 181
QY 61 VDSRDIPLGPTNPRLAAATSETCLGG-----FEVLHDKGFLDLNTQIGPSAPR 111
Db 182 IENRGYLA VPNKSLI-----LLGNIAPTSNATGKILIDGGVQVQALNG-----R 226
QY 112 VEVOADGTHAAIG-EKNGLEVSVTLSPQWSSLSQIDTEG-KRFFVFTGGRGSGHPMT 169
Db 227 VEIGLVEPGFIGNVDGNQLSLTFP-----DSVAKTDISSINNNTVFTSGAGGG--DIWV 280
QY 170 VASDIAEAR-----TRILAKLPDNNHG--RQPKVDTRSVGV-----GSASGID 212
Db 281 NADNLSLLNYGAFTGIL-----NNQNAETQAGDISINATGIVTVAQNSIISNSLIG 335
QY 213 D-----GVVSETHSTTNSVSRDPKFWVSVGAIAAGLAGLAATGIAQALALTPEDDPTT 268
Db 336 DSGKINIVAQSLQIDNSSVQS-----FSLQNSGT-----VNVKVDITVS 376
QY 269 TDPDQANAA--ESATKQQLTQEAFAKNPENQK---VNIDA-----NGNAIPSGELXDDI 317
Db 377 LLSGGISSSVRPGASASSLSQNLGLTPTGRKSGGINIQARSLLAVDSSAISASNFLADD 436
QY 318 VEQIAQAQAEQAVQAQAVENQAQOQRYEQHARRQEEQLSSGIGYGLSSALIVAGG 377
Db 437 SGDIKIQTATDAVILNRRSTISSAFQ-----GKSGNLSINTNWLNIINNSQITANT 488
QY 378 ICAG-----VTTALHRRNPQAEQTTHHTVWQQQTGG---IPQHKVALMPQERRRF 427
Db 489 LGTGAGDINI IALDINIKSLITGSTSSFLNTIIVNLGNAGNINQTARINL----- 540
QY 428 SDRRDSQGSVASTHSDSSSEVVNPNYAEVGARNLSAHQPEEHYDEVAAD--PCYSVI 485
Db 541 -----TNSGFIISTSGRPEQKET-----SGFGGNINITATELIE--IDPKGATDIITGFST- 589
QY 486 QNPSGS--GPVTGRLLIGTPGQGIQSTYALLANSGLRLGGLTSGGETAVSSVNAAPT 542
Db 590 RTFSGSRAGDIT-----LNTKNLIVRNGGVIMADAANLGNAGNININASDT 637
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:55:50 ; Search time 26.7563 Seconds
(without alignments)
3816.109 Million cell updates/sec

Title: US-09-189-415A-2

Perfect score: 2800

Sequence: 1 MPTGNLGNVNGNHLIPAP.....GETAVSSVNAAPTGPVRFV 549

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 684280 seqs, 185983659 residues

Total number of hits satisfying chosen parameters: 684280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	5.3	831	11	US-09-952-267-1
2	141.5	5.1	1488	10	US-09-738-626-5495
3	139	5.0	966	12	US-10-304-454-2
4	138.5	4.9	2353	10	US-09-797-862-33
5	137.5	4.9	1829	15	US-10-156-761-10049
6	137	4.9	2468	12	US-10-246-330-4
7	136	4.9	1569	12	US-10-287-274-312
8	136	4.9	2086	9	US-09-815-242-5639
9	136	4.9	5795	9	US-09-815-242-12610
10	136	4.9	7746	15	US-10-156-761-7965
11	135.5	4.8	1138	15	US-10-074-475-194
12	135	4.8	2344	9	US-09-815-242-12713
13	133	4.8	864	15	US-10-156-761-12878
14	132.5	4.7	1026	9	US-09-379-931-7
15	132.5	4.7	1026	12	US-10-223-597-7

16	132.5	4.7	5215	10	US-09-861-289-2	Sequence 2, Appli
17	132.5	4.7	5215	10	US-09-860-846-2	Sequence 2, Appli
18	132.5	4.7	5215	11	US-09-988-384B-2	Sequence 2, Appli
19	132.5	4.7	5215	11	US-09-836-821-2	Sequence 2, Appli
20	132.5	4.7	5215	12	US-10-371-889-45	Sequence 45, Appl
21	132.5	4.7	6146	15	US-10-156-761-10436	Sequence 10436, A
22	131.5	4.7	707	15	US-10-128-714-8571	Sequence 8571, Ap
23	130	4.6	1260	15	US-10-245-802-8	Sequence 8, Appli
24	129.5	4.6	2834	12	US-10-085-959-252	Sequence 252, App
25	127.5	4.6	771	12	US-10-205-219-175	Sequence 175, App
26	126.5	4.5	1089	12	US-10-379-632-109	Sequence 109, App
27	126.5	4.5	1209	12	US-10-379-632-96	Sequence 96, Appl
28	126.5	4.5	1684	12	US-10-238-075-877	Sequence 877, App
29	126	4.5	376	12	US-10-084-843-202	Sequence 202, App
30	126	4.5	376	12	US-10-193-002-197	Sequence 197, App
31	126	4.5	1073	15	US-10-156-761-12156	Sequence 12156, A
32	125.5	4.5	1108	12	US-09-949-029-108	Sequence 108, App
33	125	4.5	2442	14	US-10-109-886-10	Sequence 10, Appl
34	124.5	4.4	729	9	US-09-287-849-2	Sequence 2, Appli
35	124.5	4.4	729	12	US-10-359-460-2	Sequence 2, Appli
36	124.5	4.4	729	12	US-10-098-732A-16	Sequence 16, Appl
37	124.5	4.4	2514	12	US-10-320-800-40	Sequence 40, Appl
38	124	4.4	506	15	US-10-156-761-12648	Sequence 12648, A
39	124	4.4	1180	12	US-10-193-764-61	Sequence 61, Appl
40	124	4.4	1188	12	US-10-193-764-59	Sequence 59, Appl
41	124	4.4	1536	12	US-10-193-764-63	Sequence 63, Appl
42	123.5	4.4	1270	15	US-10-245-802-16	Sequence 16, Appl
43	122.5	4.4	507	9	US-09-910-087-19	Sequence 19, Appl
44	122.5	4.4	729	12	US-10-098-732A-18	Sequence 18, Appl
45	122.5	4.4	930	12	US-10-098-732A-65	Sequence 65, Appl

ALIGNMENTS

RESULT 1

US-09-952-267-1
; Sequence 1, Application US/09952267
; Publication No. US20030032772A1
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENBURG, ROSS A.
; TITLE OF INVENTION: USP22 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/952.267
; CURRENT FILING DATE: 2001-09-12
; PRIOR APPLICATION NUMBER: 09/336,447
; PRIOR FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-952-267-1

Query Match	5.3%	Score 149;	DB 11;	Length 831;
Best Local Similarity	19.2%	Pred. No. 0.0054;		
Matches	101;	Conservative	73;	Mismatches 193;
				Indels 160;
				Gaps 21;
Qy	74	SRLLAAATSETCLGFEVLHDKGFLDILNTOIGP----	SAFRVEVOADTHAAIGKNGLE	130
Db	226	SGTTGVTNSVLLGNETA--GKQATTVKNAEVLGSLTGFA	GSKAENGVSVEGGER	283
Qy	131	VSVTLSPQWSSLSQIDTEGKNRFVFTGGRGSGHPMVTVAS	----	DIABARTLILAKLD
Db	284	QIVNVGA---GQISDTSTDAVN-----	GSQHLALATVVDDNQYDILNNRADILNNQD	332
Qy	187	P-----DNHGRQPKDVT-----	-----	RSVGVS-----

Db 333 DIKDLQKEVKGIDNEVGELSRDINSLSHDVTDNQDDIKELKRGVKELDNVEGVLSRDINS 392
Qy 208 -----ASGIDGCVSETHSTTNSVSRDPKFWVSVAIAAGLAGLAATGIAQALALTP 262
Db 393 LHDDVADNODDIKAKNADKGLNKEVKELDK-----EVGLSRDIGSL----- 435
Qy 263 PDDPTTTPDQAAANAESATQDQTQEAFFKNPE-----NOKVNIDAN----- 304
Db 436 HDDVATNQADIAKNQADIKTLNNEVEELLNLSGRLDQKADIDNNINNIYELAQOQDOH 495
Qy 305 -----GNAIPSG--ELXDDIVEGIAQOAKAEAGVAVQQAQVESAQAQOQYEBQHARRQ 355
Db 496 SSDIKLKNVNEEGLDLSGRLLDQKADIAKNQADIAQNQTDIQLAAYNELQDQYAOQK 555
Qy 356 EE-----LQSSGIG--YGL-----SSALIYAGGIGAGVTTAL 386
Db 556 TEADALNKASENTONIAKNQADIANNNIYELAQOQDOHSSDIKTLAKVSAANTDRI 615
Qy 387 HRRNQPAE---QTITTTTHVVQQTGGIPQHKVALMPQERRFRDRDSQGSVAETHWS 443
Db 616 AKNKAEDASFEITLKNQNTLIEQGEALVEQNK-----INQLEGFPAAHADIQ----- 664
Qy 444 DSSSEVNPYAEVGGARNSLSAHOPEEHYDEVAADPGYSVIONPSG 490
Db 665 --DKQILQONQADITTKNTAI-----EQNINRTVA--NGFEIEKNKAG 702

RESULT 2

US-09-738-626-5495
; Sequence 5495, Application US/09738626
; Publication No. US20020197605A1

GENERAL INFORMATION:

; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1998-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 5495
; LENGTH: 1488
; TYPE: PRT
; ORGANISM: Corynebacterium glutamicum

US-09-738-626-5495

Query Match 5.1%; Score 141.5; DB 10; Length 1488;
Best Local Similarity 18.9%; Pred. No. 0.049;
Matches 106; Conservative 75; Mismatches 225; Indels 155; Gaps 19;

Qy 48 SLF-SPLRNSMADSVDRDIPGLPTNPSRLAAATSETCLLGGFEVLHDKGPLDILNTQIG 106
Db 778 ALFEQLADLSNESAELVAGAGMGLVGMALATSGNWAIGMVI----- 822
Qy 107 PSAFRVEQADGTHAAIGKNGLEVSVTLSPQWSSLOS-ID---TEGKNRVFT----- 157
Db 823 -----TVASSMGVIIFTIIMKVRGSLVSGVDEAVTSTVINRFLDTQVSSA 867

Qy 158 GRRGGSGHPMWTVASDIAEARTILAKLDPDNHGGROPKVDTRSVGVGSASGIDDDGWS 217
Db 868 GATSGDGMERRAAATGLIGATHMVLNRDGDGGG-----SDSGSGSGSGSDSLGGEKAAG 923
Qy 218 ETHSTTNSVSRDPKFWVSVAIAAGLAGLAATGIAQALALTPEDDPTTTPDQOANA 277
Db 924 LAKVVT-----VAGAGLVGKYATD-----ALDNYADGVINGDGDGAPAA 963
Qy 278 AESATKDQLTQEAFFKNPENQKVNIDANGNAIPSGELXDDIVEGIAQOAKAE----- 328
Db 964 GGDATVDG-----DYVADGDAIASADANADFDVGVVDGAGRASFNSAAYSS 1009
Qy 329 -GEVARQAVESNAQAQRYEDQHARQEBELQSSGIGYGLSSALIIVAGGI-GAGVTTAL 386
Db 1010 DGTLLDCEGASVDAQGNPLHADGTPMSAAEAEMKNA-GLSSSGTMMKSGVKSSGITTA 1068
Qy 387 HRRNQPAEQITTTTHVVQQTGGIPQHKVALMPQERRFRDRDSQGSVAST----- 440
Db 1069 DVMDQDQSLASVTSGLSKIPTVY-----ADVSGAGTIVGTTGADYSA 1112
Qy 441 -----HWSDSSEVNPYAEVGGARNSLSAHOPEEHYDEVAADPGYSVIONPSG-- 491
Db 1113 TDSSAGLNMSEALQSGTPMGALAGGSVS-SSDOAMNDAAALQIAASQGLAPAGSIAGMEQ 1171
Qy 492 -----GPTVTRLIGTP-----GGIQTSTYALLANSGLRLGMGL 526
Db 1172 LSAQATEAPAGKAGKQLGDLSSGALSALNTQLASMGQGVGDSVNSAYA-----AG---GMGGV 1223
Qy 527 TSGGETAVSSVNAAPTGPVR 547
Db 1224 DVAGKVTEANAQHLSQLVPGQIQ 1244

RESULT 3

US-10-304-454-2
; Sequence 2, Application US/10304454
; Publication No. US20030200567A1

GENERAL INFORMATION:

; APPLICANT: Komatsu, Setsuko
; APPLICANT: Sharma, Arun
; APPLICANT: Hashimoto, Junji
; APPLICANT: Sakaguchi, Kengo
; TITLE OF INVENTION: Cold Stress-Responsive CRTintP Gene And
; FILE REFERENCE: 3462.1001-000
; CURRENT APPLICATION NUMBER: US/10/304, 454
; CURRENT FILING DATE: 2002-11-25
; PRIOR APPLICATION NUMBER: JP 2002-121275
; PRIOR FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 966
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-304-454-2

Query Match 5.0%; Score 139; DB 12; Length 966;

Best Local Similarity 20.9%; Pred. No. 0.041;
Matches 128; Conservative 82; Mismatches 223; Indels 178; Gaps 31;

Qy 23 PSQTDGAARGGTGHLISSTGALGSRSLFSP-LRNSMADSVDRDIPGLP-----TNP 73
Db 327 PSPADAFNAGSSVYINSAGP--NPIMVQPSFQNTPPGVSS--IPVLGGISGAFGIVDP 382
Qy 74 SRLAAA-TSETCLLGGFEV-----LHDKGPLDILNTQIG-PSAFRVE-VQADGTHAA 122
Db 383 SRTSAVNTHGTSTTSGSAGMTTASAGAVNEGROVVERTQGCNPSATSMHGLPARTVIAA 442
Qy 123 IGEKNGL-----VSVTLSPQWSSLS-----QSIDTEGKNRVFTTCGRSGHPMTVA-- 171
Db 443 IPARSTAEAPNHLVLSVILFVQVRQVAMPNQSTVSGSQTA-----GGGQSPQASVGGV 497


```
QY 172 ----SDIAEAKTRILAKLDPNHHG-----GRQPKVDVTRSVGVGSAGSID--DGWVSETH 220
Db 498 ASIPSIQAQVTAQVANAALSANQQGVSSSAQNTVDQGRSV---TTNGVDNVDSLVS--- 551
QY 221 TSTTSSVRSOPKFWVSGATAAGLAGLAATGIAQALALTPPPDDPTTTDDPQAANAES 280
Db 552 ASTQLQNELSDSN-----NGRTSLNAQSLVAGAGISPS-----NTSDPN-----L 591
QY 281 ATKDQLTOEA-----FRNPENQKNID-----ANGNAIPSGELXDDIVEQIAQQA 325
Db 592 ASEDSTENAPNIGSIQHPMEGIGHADNRKPSSEITANLV--GQITTTCTDDIS--V 647
QY 326 KEAGEVARQAVESNAQAQRYEDQHARRQEBELQLSSGI-----GYG 367
Db 648 NRSAESQKNIPLDGVSAQSIKPSASSRSEPVGLGGGLQPKRRSRRTAKPPGSSSDTGEV 707
QY 368 LSSALIV---AGGIGAGVTTALHRRNQPAEOT---TTTTTHVVQOQTGGIPQHKVALMP 421
Db 708 VNSSRISNSQNAVSGQVQLQALASQNTNVNRKSHVTDSPLPSTTSQFSGGMPFRR----- 762
QY 422 QERRRFSRRDSQGSVASTHWSSESSEVNPYA-----EVGGARNSLSAH 466
Db 763 -----QGEQGVDFGSMSSVLNPNPFCGNLLSNVAEQTGMGSAGDLRNWV--- 807
QY 467 QPEEHYDEVAADPGYSVIONFSGG-----PVTGRLIGTPGQGIQSTY 510
Db 808 --EECAQSPAIMDTMSNLVQNVDSGRGQGGIDLSRMQOMPVVSVQLGGAGARPACT- 864
QY 511 ALLANSGLRL 521
Db 865 ----NSGQSRL 871
```

RESULT 4

```
US-09-797-862-33
; Sequence 33, Application US/09797862
; Patent No. US20020102276A1
; GENERAL INFORMATION:
; APPLICANT: PEAK, IAN RICHARD ANSELM
; APPLICANT: JENNINGS, MICHAEL PAUL
; APPLICANT: MOXON, E. RICHARD
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0134
; CURRENT APPLICATION NUMBER: US/09/797,862
; CURRENT FILING DATE: 2001-05-03
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-797-862-33
```

```
Query Match 4.9%; Score 138.5; DB 10; Length 2353;
Best Local Similarity 22.0%; Pred. No. 0.17;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;

QY 6 LQNNVNGNHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVSRD 65
Db 1163 LGGTANG-----RNDTGVINKDG-----LITILANGAAAGTDSAN 1198
QY 66 IGPLTPNPSRLAAATSETCLLGGPEVLHDKGPL-----DILNTQIGPSAFRVEVQADGTHA 121
Db 1199 -----GNTISVTKDGISAGNKEITNVKALKTKYKDTQNTI-----ADETDKEPHA 1243
QY 122 AIGENKNGLEVSVTLSPQEWSSLSQSDTEGKRRFVFTGCGGSGHPMTVVASDIAEAPRI 178
Db 1244 AV--KNANEVEFV-----GKNGATVSAKTDNNGKHTVTI--DVAEAKVGD 1284
```

```
QY 179 ----TRILAKLDPNHHGRQPKVDVTRSVGVGSAGSIDDGCVSVBETHST-TNSSVR--- 229
Db 1285 GLEKDTGKIKLKVDTNDGNNLLTVDA--TKGASVAKGEFNAVTTDATTAQGTNANERKGV 1343
QY 230 -----SPKFWVSGATAAGLAGLAATGIAQALALTPPPDDPTTTDDPQAANA 277
Db 1344 VVKGSGNATATETDKKVKATVGDVAKAIND--AATFV-----KVENDDSATIDDSPTDDG 1396
QY 278 AESATK--DOLTOBAFKNPENQKNID-----ANGNAIPSGELXDDIVEQIAQQA 327
Db 1397 ANDALKAGDTLTLKAGK---NLKVRDGNKITPALANDLSVKSATVSKL-----S 1444
QY 328 AGEVARQAVESNAQAQRYEDQHARRQEBELQLSSGIYGLSSALIYAGGIGAGVTTALH 387
Db 1445 LGTNGKNVITSDTKGLNFAKDSKTGDDANIHL--NGIASTLTDTL-----LNSGATTNL- 1497
QY 388 RRNQPAEQTTTTTHVVQOQTGGIPQHKVALMPQERRRFSRRDSQGSVASTHWSDES 447
Db 1498 -----GG-----NGITNEKKRAASVKD----- 1515
QY 448 EVNPNYAEVGGARNSLSAHQPEEHYDEVAADPGYSVIONFSGGSPVTGRLIGTPGQGIQ 507
Db 1516 -VLNAGNVNRGVKA--SANNQVENI--DEVAT---YDVTDFVSGDKDTSVTVESKDNCKR 1569
QY 508 STYALLANSGLRLGMLGTSGGE 531
Db 1570 TEVKIGAKTSVIKDHNGKLFITKE 1593
```

RESULT 5

```
US-10-156-761-10049
; Sequence 10049, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10049
; LENGTH: 1829
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10049
```

```
Query Match 4.9%; Score 137.5; DB 15; Length 1829;
Best Local Similarity 20.9%; Pred. No. 0.14;
Matches 126; Conservative 84; Mismatches 257; Indels 135; Gaps 23;

QY 5 NLGNVNG--NHLIPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSV 62
Db 676 DLTSVNGMAGNLTAQVRNIAQVATAVARGDLSQKID---VDARGEILELKNLTNTWVD 731
QY 63 SRDIPGLTPNPSRLAAATSETCLLGG--FEVLHDKGPLDILNTQIGPSAFRVEVQADGTHA 121
Db 732 --QLSNFAEQVTRVAREVTEGMLGQAEVQGVSGTWKDLTQSVNFMANNLTIOV----- 784
QY 122 AIGENKNGLEVSVTLSPQEWSSLSQSDTEGKRRFVFTGCGGSGHPMTVVASDIAEAPRI 181
Db 785 ----RNTAEVTTAVAKGDLSSKITTADAKGEILELVT-----TVNTWVDQLSSFAEQVTR 835
QY 182 LAKLDPNHHGRQP-----KD-----VDTRSVGVGSAGSIDDGWVSET 219
```

Db 836 AREVTEGILGQAHVPGVTGIWKDLSNNVNMANNLTQVRNIS-QVAAAVANGDLTRT 894
Qy 220 HT-----STTNSVSRSDPKFVSVGAIAA-----GLAGLAA--TGIAQALALTP 261
Db 895 VTIEARGEVAQLADFTNTWKTLSFADQVTKVAREVGTGDLGILGQAHVPGVA----- 947
Qy 262 EPDDPTTDPQAAANAASATKQ-----LTQEAFKPENQKVNIDANGNAIPSGELXD 315
Db 948 -----GTWKDLTSVNGMASNLTGQVRNIAMVTTALAKGDLTKKIDIDARGEILELKTINN 1003
Qy 316 DIVEQIAQQAKEAGEVARQAVESNAQAQORVEDQHARRQBELQLSSGIGYGLSSALIVA 375
Db 1004 TMDQLSSFAEVEVTRVAREVGTGQLGGQARVRDVGDTWRDLTESVNEAGNLTKQVRAI 1063
Qy 376 GGIGAGVTTA-LHRR-----NQPAAEQTTTTTHTTWWQQQTGGIPQKVA 418
Db 1064 ARVATVTRGDLNLKIDVDASGEIQELODYINKMIANLRDITIANKEQDLKGNLARISA 1123
Qy 419 LMPQERRFRDRRDSQGSVASTHWSSEVVNPNY-----AEVGGARNSLSAHQ 467
Db 1124 LM-QGRRLDD-----VASLIMSELTVPVSAQHGAPFLAMPLVDAREGGAE----- 1168
Qy 468 PEEHIYD-EVADPGYSVIONFSGSPVTR-----LIGTPGQIGQISTVALLANSGLRLG 522
Db 1169 -QEDAYELRMCGSYGYSM-----GSMPTSFREGEALIGTAAQEKRTILVENAPSGYLKIS 1222
Qy 523 MG 524
Db 1223 SG 1224

RESULT 6

US-10-246-330-4
; Sequence 4, Application US/10246330
; Publication No. US20030166030A1
; GENERAL INFORMATION:
; APPLICANT: O'Toole, George A.
; APPLICANT: Mah, Thien-Fah
; TITLE OF INVENTION: METHODS TO STUDY AND MECHANISMS OF
; TITLE OF INVENTION: BIOFILM-SPECIFIC ANTIBIOTIC RESISTANCE
; FILE REFERENCE: 14537-002001
; CURRENT APPLICATION NUMBER: US/10/246,330
; PRIOR FILING DATE: 2002-09-18
; PRIOR APPLICATION NUMBER: US 60/323,241
; PRIOR FILING DATE: 2001-09-18
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 2468
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-10-246-330-4

Query Match 4.9%; Score 137; DB 12; Length 2468;
Best Local Similarity 21.5%; Pred. No. 0.23;
Matches 135; Conservative 57; Mismatches 275; Indels 160; Gaps 28;

Qy 2 PIGNLGNVNGNHLIPAPPLPSQT-----DGAARGGTGHLISSTGALGSRSLFSLPLR 54
Db 455 PIGQTSADANGNWSFTFGSQLPDGTVVNVVARDAAAGNSSPATSIYVDGVAPNAPVVEPSN 514
Qy 55 NS-----MADSVSRDIPGLTPNPSRLAA-----ATSETCLGGFEVLHDKGPLDILN 102
Db 515 GSELSTAEFGSSVTLTDGNGNPIGQTTADANGNWSFTSTPLPDGTVV--NVVARDAG 572
Qy 103 TQIGSAFREVQADGTHAAIEKNGLEV-----SVTLSPQEWSSLSQSIDTECKNRP 154
Db 573 NSSPPASVTVDAVAPAT-PTVDPNSNGTTLSGTAEPGSSVTLTDGNGNPIGQTTADGSGNW 631
Qy 155 VFTGGR-----GSGGHPMTVVASDIAEAETILAKLDPN-----HGGROP 195
Db 632 TFTPSTPLPNGTVVNNATATDPSGNASSPASVTVDAVAPATPV---VNPSNGTTLTGTAEP 688

Qy 196 KDVTTRSVGVGSAGS--IDGCVVSEHTSTT---NSSVSRSDPKFVSVGAIAAGLAGLAA 250
Db 689 GATVTLTDGNGNPIGQTTADGSGNWSFTPTTLPNGTV-----VNATATDASGNTS 739
Qy 251 TG-----IAQALALTP--EPDDPTTDPQAAANAASATKDQLTOEAFKPNQKVNIDAN 304
Db 740 AGSVTVDSVAPATPVINPSNGT---LSGTAEPGSSVTLT-DGNGNPIGQ-VTADGS 792
Qy 305 GN--AIPSGELXDDIVBQIAQAQAGEVARQAVESNAQAQORVEDQHARRQBELQLSS 362
Db 793 GNSFTSTPLADGTVVN-ATATDPAGNTSQGGSTT-----VDGVAPTTPTVNLSN 842
Qy 363 GI-----GYGLSALIVAGGIGAGVTALHRRNQPAEQTTTTTHTTWWQ 406
Db 843 GSSLGTAEPGSTVILTDGNGNPIAEVTADGSGNWTYT-----PSTPIANGTVVNVVA 895
Qy 407 QQTGGIIPQHVALMPQERRFRDRRDSQGSVASTHWSSEVVNPNYAEVGGARNSLSAH 466
Db 896 QDAAG-----NSPGASVTVDSQAPAPVNP-----SNGTTLTG- 930
Qy 467 QPEEHIYDEVAADPGYSVIONFSGSPV-----TGRLLGTGQIGQISTVALLANSGL 519
Db 931 -----TAEFGATVTLTDGNGNPIGQTTADGSGNWSFTPTCTPLANGTVVNTASD- 979
Qy 520 RLGMGLTSGGETAVSSV-NAAPTPGP 545
Db 980 --PTGNTSAPASTTVDSVAPAPVNP 1004

RESULT 7

US-10-287-274-312
; Sequence 312, Application US/10287274
; Publication No. US20030181408A1
; GENERAL INFORMATION:
; APPLICANT: Forsyth, R. Allyn
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; TITLE OF INVENTION: GENES ESSENTIAL FOR MICROBIAL PROLIFERATION AND ANTISENSE THERETC
; FILE REFERENCE: ELITRA.008DVI
; CURRENT APPLICATION NUMBER: US/10/287,274
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: US 60/164415
; PRIOR FILING DATE: 1999-11-09
; PRIOR APPLICATION NUMBER: US 09/711164
; PRIOR FILING DATE: 2000-11-09
; NUMBER OF SEQ ID NOS: 469
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 312
; LENGTH: 1569
; TYPE: PRT
; ORGANISM: Escherichia coli
US-10-287-274-312

Query Match 4.9%; Score 136; DB 12; Length 1569;
Best Local Similarity 20.9%; Pred. No. 0.15;
Matches 127; Conservative 69; Mismatches 259; Indels 154; Gaps 26;

Qy 19 APPLPSTQTDGAARGT--GHLISSTGALGSRSLFSLPNSMADSVDSRDIPLCLPTNPSRL 76
Db 579 AKHVEQQGGGALIASTTSGTLLTGNTSVGDADF---INNSEAKV-----VLENAGSL 628
Qy 77 AAATSETCLGGFEVLHDKGPLDILNTOIGP---SAFREVOADGTHAAIEKNGLEVSV 133
Db 629 TVVTGSAV---DTIINANGKMDVVGKDVGVLSNAGTQTIYASATSDKANIKGKGQTVY 685
Qy 134 TILSPQE--WSSLQSIDTECKNRFVFTGCGSGGHPMTVVASDIAEAETILAKLDPD--- 188
Db 686 GLATEANIESGEQIVDGGSTKTHINGGTQTVQNGKAINTDIVSGLQOIWANGTAEGSI 745
Qy 189 -----NHGRQPKDV-----DTRSVGVGASGI---DDGVSVSETHSTTTNSSVRS 230
Db 746 INGSQVNVNEGGLAENSVLNDGGTLDVRE--KGSATGIQQSSQSGALVATRAVTRVGTGA 803

;; PRIOR APPLICATION NUMBER: JP 2001-204089
;; PRIOR FILING DATE: 2001-05-30
;; PRIOR APPLICATION NUMBER: JP 2001-272697
;; PRIOR FILING DATE: 2001-08-02
;; NUMBER OF SEQ ID NOS: 15109
;; SEQ ID NO 12878
;; LENGTH: 864
;; TYPE: PRT
;; ORGANISM: Streptomyces avermitilis
US-10-156-761-12878

Query Match 4.8%; Score 133; DB 15; Length 864;
Best Local Similarity 21.5%; Pred. No. 0.11;
Matches 124; Conservative 84; Mismatches 191; Indels 178; Gaps 32;
QY 5 NLGNVNGNHLPPAPPLPSQTDGAAGGTGHLISSTGALG-SRSLFPLRNSMADSV- 62
DB 353 NFNAPAGAN-----PPALSALPGL--GSGGL-SQLSGLIGASLDPPGFARFDSAPV 403
QY 63 SRD--IPGLPTNPRLAAATSETCLLGGFEVLHDKGP---LDILNTOIGPSAFRVEQAD 117
DB 404 TFDLRISSPATVHKVSTEDTVLFG--KYVDVPGGTQKVLPSQL-VTPIRVE--- 455
QY 118 GTHAAGKNGLEVSVTLSPOWSSLSQIDTEGKRFVFTGGRGSGHPMTVVASDIAEA 177
DB 456 -----GAKEGKDVTL-----SLPAIDHEVR-----SGH----- 478
QY 178 RTRILAKLDPNHGKQPKVDTRSVGSGASGIDGVVSEHTSTTNSVRSO----- 231
DB 479 LRLV-----LASTDLGYASPTAPATYTVSLUKDULLVPTA 513
QY 232 PKFWSVGVAIGAAGLAGAATGIAQALAL-----TPEDDPTTDP-----DQANAA 278
DB 514 PGVTTAAAPLPAWVWPLAGVVALALLSGRRRTATPAPDPAALAEVPLEITDLSKRYA 573
QY 279 ESATKDQLTQEAFFKPNQKVN-----DAN-----GNAIPSGE 312
DB 574 KSADRYAVRDLFRVEKGQVLGLLPGNGAGKTTTLRLMLGLIKPDAGEIHVFGHAIRPGA 633
QY 313 LXDIVEQIAQAQAKAG---EVARQAVESNAQAQR-YEDOHARRQELQLSSGIGYGL 368
DB 634 ----PVLRSVGSFVEGAGFLPHUSGRENLELYWQATGRPSED AHL--DEALQI-AGLGAL 687
QY 369 SSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHTVVOQQTGGIPQHKVAMPQERRRFS 428
DB 688 ARAV-----RTYSGMQRALIAQAMGLPDLILDEPTNGLDPPQI---REMREVM 736
QY 429 DRDSQGS--VASTHW-----SDSSSEV-----NPYAEVCGARNSL---SAHQPE 469
DB 737 IRYAAGRVTIVVSSHLLAEVEQSCTHLVYMDRGLVQAGFVHEIVGSGDTLLVGTGTPVD 796
QY 470 EHYDEVAADPGV-SVIQNFSG-----SGPVTGRLI 499
DB 797 EPVKEKVGALPGVASAVRTDEGLLVRLDLDGSGAGRLV 833

RESULT 14

US-09-379-931-7
;; Sequence 7, Application US/09379931
;; Patent No. US20020009792A1
;; GENERAL INFORMATION:
;; APPLICANT: Smit, John
;; APPLICANT: Bingle, Wade H.
;; APPLICANT: No. US20020009792A1
;; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAUL
;; FILE REFERENCE: 08106/002003
;; CURRENT APPLICATION NUMBER: US/09/379,931
;; CURRENT FILING DATE: 1999-08-24
;; PRIOR APPLICATION NUMBER: US 08/614,377
;; PRIOR FILING DATE: 1996-03-12
;; PRIOR APPLICATION NUMBER: US 08/194,290
;; PRIOR FILING DATE: 1994-02-09
;; PRIOR APPLICATION NUMBER: US 07/895,367

;; PRIOR FILING DATE: 1992-06-09
;; NUMBER OF SEQ ID NOS: 12
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 7
;; LENGTH: 1026
;; TYPE: PRT
;; ORGANISM: Caulobacter crescentus
US-09-379-931-7

Query Match 4.7%; Score 132.5; DB 9; Length 1026;
Best Local Similarity 21.8%; Pred. No. 0.15;
Matches 123; Conservative 64; Mismatches 222; Indels 155; Gaps 27;
QY 26 TDGAAGGTGHLISSTGALGSRSLFPLRNSMADSVDRIPGLPTNPRLAAATSETCL 85
DB 230 TDNAAGVNLFTAYPSGVSSTL-----SLTTGTD--LTGTANNDTFVAGEVAGAA 280
QY 86 LGGFEVLHDKGPDIILN-----TOIGPSAFRVEQADGTHAAIGCKNGLEVSVT 134
DB 281 LTVGDTLSSGAGTDVLNWWQAAAVTALPTGVTISGIEETMNTVSGAAILTNTSSGVTGLTA 340
QY 135 LSPQWSSLSQISIDT-EGKRFVFTGGRGGS-----GHPMTVVASDIAEAARTRILAKLPD 188
DB 341 LNTNTSAAQTVTAGAQNLTATTAQAANNVAVDGRANVTVAS-----TGVT----- 389
QY 189 NHGGRQPKVDTRSVGSGASGIDGVVSEHTSTTNSVRSRDPKFWVSVGAIAAGLAGL 248
DB 390 -----GTTTVGANSAA---SGTVS---VSVANSST-----TTTGAIA--VTGG 424
QY 249 AATGIAQALALTPEPDDPTTDPQANAAESATKQOLTQEAFFKPNQKVNIDANGNAI 308
DB 425 TAVTVAGTAG---NAVNTLTQADVTVTNGNSTTAVTVTQTA-----AATAGAT 470
QY 309 PSGBLXD--DIVEGIAQAQAKEAGEVAR-----QOAVESNAQAQRVEDQHARRQEBELQ 359
DB 471 VAGVNGAVTITSAAASATAGIAVTILGSGAAITDSSALTNNL----- 518
QY 360 LSSGIGYGLSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHTVVOQQTGGIPQHKVAL 419
DB 519 --SGTGTSL-----GIGRGALTA-----TPTANTLTILNVLGLT--TTGAIITDSEAA- 560
QY 420 MPQERRPDRDSQGSVASTHWSDSSEVSNVPAEV-----GGARNSLSAHQPEEHYDE 475
DB 561 -----ADGFTTINIAGSTASTIASLVAADATLTINISGDARVTITSH-----T 604
QY 476 VAADPGYSVIQNFSS---GSGPVTGRLLIGTFCQGIQS-----TYALLANS-GL 519
DB 605 AALITGITVNSVATLGAELATG-LVFTGAGRDSILLGATTKAIVMGAGDDTFTVTSSA 663
QY 520 RLGMGGLTSGG---ETAVSSVNAA 540
DB 664 TLGAGGSVNGGDTDLVLANVNGS 687

RESULT 15

US-10-223-597-7
;; Sequence 7, Application US/10223597
;; Publication No. US20030135037A1
;; GENERAL INFORMATION:
;; APPLICANT: Smit, John
;; APPLICANT: Bingle, Wade H.
;; APPLICANT: No. US20030135037A1
;; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULOR
;; FILE REFERENCE: 08106/002003
;; CURRENT APPLICATION NUMBER: US/10/223,597
;; CURRENT FILING DATE: 2002-08-19
;; PRIOR APPLICATION NUMBER: US/09/379,931
;; PRIOR FILING DATE: 1999-08-24
;; PRIOR APPLICATION NUMBER: US 08/614,377
;; PRIOR FILING DATE: 1996-03-12
;; PRIOR APPLICATION NUMBER: US 08/194,290
;; PRIOR FILING DATE: 1994-02-09
;; PRIOR APPLICATION NUMBER: US 07/895,367

Search completed: December 16, 2003, 09:07:33
Job time : 28.7563 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 25.7653 Seconds
(without alignments)
901.547 Million cell updates/sec

Title: US-09-189-415A-2

Perfect score: 2800

Sequence: 1 MPIGNLGNVNGNHLIPAP.....GETAVSSVNAAPTEGPRFV 549

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	149	5.3	831	4	US-09-336-447A-1
2	144	5.1	3892	4	US-09-328-352-5503
3	141	5.0	580	2	US-08-672-564-1
4	139.5	5.0	2354	4	US-09-268-347-47
5	139.5	5.0	2411	4	US-09-268-347-36
6	138.5	4.9	2353	3	US-09-377-155-33
7	138.5	4.9	2353	3	US-08-913-942-4
8	138.5	4.9	2353	4	US-09-669-974-33
9	137	4.9	2736	4	US-09-252-991A-30227
10	134.5	4.8	734	4	US-09-328-352-4412
11	134.5	4.8	1912	1	US-08-409-985-4
12	134.5	4.8	1912	3	US-08-685-467-4
13	134	4.8	375	4	US-09-252-991A-21021
14	133	4.8	2703	1	US-08-185-432-19
15	133	4.8	2703	4	US-08-899-232-4
16	132.5	4.7	1026	1	US-08-194-290-7
17	132.5	4.7	1026	2	US-08-614-377A-7
18	132.5	4.7	1026	3	US-09-142-648B-7
19	132.5	4.7	5215	3	US-09-105-537-2
20	132	4.7	1415	4	US-09-252-991A-26438
21	128	4.6	1507	6	5268270-2
22	126	4.5	376	4	US-09-056-556-202
23	126	4.5	376	4	US-09-072-596-197
24	126	4.5	444	4	US-09-252-991A-21658
25	126	4.5	646	4	US-09-252-991A-23299
26	125	4.5	851	4	US-09-252-991A-22021
27	125	4.5	2442	4	US-09-514-247A-10

28	124.5	4.4	729	4	US-09-223-040-2	Sequence 2, Appli
29	124.5	4.4	1034	4	US-09-252-991A-26658	Sequence 26658, A
30	124	4.4	925	4	US-09-252-991A-27057	Sequence 27057, A
31	124	4.4	1180	4	US-09-206-942-65	Sequence 65, Appl
32	124	4.4	1188	4	US-09-206-942-63	Sequence 63, Appl
33	124	4.4	1536	4	US-09-206-942-67	Sequence 67, Appl
34	123.5	4.4	665	4	US-09-252-991A-20383	Sequence 20383, A
35	122.5	4.4	507	3	US-08-860-635A-19	Sequence 19, Appl
36	122.5	4.4	507	4	US-09-281-476-19	Sequence 19, Appl
37	122	4.4	2137	4	US-09-134-001C-4463	Sequence 4463, Ap
38	121	4.3	608	4	US-08-843-572E-2	Sequence 2, Appli
39	121	4.3	1638	4	US-09-071-035-258	Sequence 258, App
40	121	4.3	1638	4	US-09-071-035-262	Sequence 262, App
41	121	4.3	1638	4	US-09-071-035-266	Sequence 266, App
42	120.5	4.3	2039	4	US-09-077-098A-7	Sequence 7, Appli
43	120	4.3	2042	4	US-09-077-098A-6	Sequence 6, Appli
44	119.5	4.3	729	4	US-09-625-188-20	Sequence 20, Appl
45	118.5	4.2	1287	4	US-09-252-991A-29606	Sequence 29606, A

ALIGNMENTS

RESULT 1
US-09-336-447A-1
; Sequence 1, Application US/09336447A
; Patent No. 6310190
; GENERAL INFORMATION:
; APPLICANT: HANSEN, ERIC J.
; APPLICANT: AEBI, CHRISTOPH
; APPLICANT: COPE, LESLIE D.
; APPLICANT: MACIVER, ISOBEL
; APPLICANT: FISKE, MICHAEL J.
; APPLICANT: FREDENEURG, ROSS A.
; TITLE OF INVENTION: USP21 AND USP22 ANTIGENS OF MORAXELLA CATARRHALIS
; FILE REFERENCE: AMCY:024
; CURRENT APPLICATION NUMBER: US/09/336.447A
; CURRENT FILING DATE: 1999-06-21
; NUMBER OF SEQ ID NOS: 98
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Moraxella catarrhalis
US-09-336-447A-1

Query Match	5.3%	Score 149;	DB 4;	Length 831;
Best Local Similarity	19.2%	Pred. No. 0.00027;		
Matches 101;	Conservative 73;	Mismatches 193;	Indels 160;	Gaps 21;
Qy	74	SRLAAATSETCLGGFEVLHDKGFLDILNTOIGP---SAPRVEVQADGTHAIGEKNGLE	130	
Db	226	SGTTGVTNSVLLGNETA--GKQATTNKAIEVGLSLTGFAGESKAENGVSVEGGER	283	
Qy	131	VSVTLSPEQWSSLOSIDTEGKRNREFTGCGSGSGHPMVTVAS----DIAEARTRILAKLD	186	
Db	284	QIVNVGA---QGISDTSTDAVN-----GSQHALATVDDNQYDINNADILNNQD	332	
Qy	187	P-----DNHGGRPQKVDVT-----RSVGVGS-----	207	
Db	333	DIKDLQKEVKGLDNEVGELSRLDINSLHDVTDNQDDIKELKRGVKELDNEVGVLSDINS	392	
Qy	208	-----ASGIDGVVSETHSTTSSVRSDPRFVSVGAIAAGLAGLAATGIAQALATPE	262	
Db	393	LHDDVADNQDDIAKNKADIKGLNKEVKELDK---EVGVLSRDIGSL-----	435	
Qy	263	PDDPTTTTDPQAAANAESATKQDLTQFAFKNPE----NQKNVIDAN-----	304	
Db	436	HDDVATNQADIAKNQADIKTLNNEVEELLNSGRLDDQKADIDNNINNIYELAQOQDOH	495	
Qy	305	-----GNATPSG--ELXDDIVFQIAQQAKEAGEVAFQQAQVESAQAQRYEDQHARRQ	355	
Db	496	SSDIKTLKNVVEGLDLSGRLIDQKADIKNQADIAQNTDIDQLAAYNBLQDYAQKQ	555	

QY 356 EE-----LQSSGIG--YGL-----SSALIVAGGIGAGVTAL 386
DB 556 TEADALNKASSENTONAKQADIANNNIYELAQOQDQSHSDIKTLAKVSAANTDRI 615
QY 387 HRRNOPAE---QTTTTHVVOQTGIGPOHKVALMPQERRRFSRRDSQGSVASTHWS 443
DB 616 AKNKAEDASPETLTKNQNTLIEQCEALVEQNK--INQLEGFPAAHADIQ----- 664
QY 444 DSSSEVNPYAEVGGARNSLSAHQBEHHYDEVAADPGYSVIONPSG 490
DB 665 --DKQILQOQADITTKNTAI-----EQNINRTVA--NGFEIEKNKAG 702

RESULT 2

US-09-328-352-5503
; Sequence 5503, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5503
; LENGTH: 3892
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5503

Query Match 5.1%; Score 144; DB 4; Length 3892;
Best Local Similarity 19.6%; Pred. No. 0.0082;
Matches 113; Conservative 80; Mismatches 253; Indels 130; Gaps 21;

QY 3 IGNLGNVNGNHLIPAPLPQSDTGAARG-----GTGHLISGTGALGSRSLFSPLRNSMA 58
DB 178 IAAASHNDSNDSSAPADTPPTSDGVTFSDVPTSDVINASEAGNVITGVKNIPA 237
QY 59 DSVDSDRIPLPTNPSRLAAATSETCLGGFVLHDKGLDILNTQIGPSAPRVEVDAG 118
DB 238 D-----AANTAVTVINGV-----LYSIDTEGKNRFVFTGGG-----GSGHP 166
QY 119 THAAIGEKNGLEVSFVTLSPQWSS-----LQSIDTEGKNRFVFTGGG-----GSGHP 166
DB 271 SGLVADADKTTIDAKVFTTDAAGNSSVNDTQYTTLDITAPNAPVIDPVNGTDPITGTAP 330
QY 167 MVTVASDIAEARTRILAKLDPNHHGGRQPKDVTDRSVGSGASGIDDDGVVSEHTSTNS 226
DB 331 GSTVTVTPYDGGSTKVV-AGPDG-----TWTV---PNPGLNDG-----DEVT 368
QY 227 SVRSDDPKFVWSGAIAGLAGLAATGIAQALALTPEDPTTDDPDQANAESATKQOL 286
DB 369 AVATDP-----AGNTSGPATAVDVAFTVALDDVLND-----STPALTGTVNDP 414
QY 287 TOEAFKNPENQKNVDANGNAIPSGELXDDIVEQIAQ-----QAKEAGEVARQQA 338
DB 415 TATVVVNVDPG--VDYPAVNVNGGTWTLAONTLPTLADGPHITVTATDAAGNVGDTGVV 472
QY 339 SNAQAQORVEDQHARQEEQLQSSGIGYGLSSA-----LIVAGGIGAGVTTLHRRNQPAE 394
DB 473 TVDTAPN-----TAGVTFIDSVDADNVINASEAGNVITGVKNIPAD 518
QY 395 QTTTTHVVOQTGIGPOHKVALMPQERRRFSRRDSQGSVASTHWSSEVVNPYA 454
DB 519 ATNTAVTVI-----NGVTYN--ATVDKTAGTWTWSVPSGSLVAD--ADKTDIAKVTFT 568
QY 455 EVCGARNSLSAHQBEHHYDEVAADPGYSVIONPSGFPVGTGLTGPCGQIQSTV---- 510
DB 569 DAAGNSSTVN-----DTQIYTLDTAAAPAVPIDVNGTDPITG--TAEPGSTVTVTPNGD 622
QY 511 ---ALLANSGLRLGMLTSGGETAVSSVNAAPT 543

DB 623 TATVAGPDGWSVPNPNGLNDGDEVEAIAITDPAGNP 658

RESULT 3

US-08-672-564-1
; Sequence 1, Application US/08672564
; Patent No. 5824503
; GENERAL INFORMATION:
; APPLICANT: KURUME, Yoko
; APPLICANT: IZU, Hiroyuki
; APPLICANT: IZUMI, Yoshiya
; APPLICANT: SANO, Mutsumi
; APPLICANT: KATO, Ikunoshin
; APPLICANT: ITO, Makoto
; TITLE OF INVENTION: GENE ENCODING ENDOGLYCOCERAMIDASE ACTIVATOR
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
; STREET: P.O. Box 747
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22040-0747
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA: US/08/672,564
; APPLICATION NUMBER: US/08/672,564
; FILING DATE: 28 JUNE 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: WEINER, Marc S.
; REGISTRATION NUMBER: 32,181
; REFERENCE/DOCKET NUMBER: 1422-0263P
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 580
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-672-564-1

Query Match 5.0%; Score 141; DB 2; Length 580;
Best Local Similarity 23.9%; Pred. No. 0.00076;
Matches 84; Conservative 45; Mismatches 126; Indels 96; Gaps 16;

QY 4 GNLGNVNGNHLIP-----APLPQSDTGAARGGTGHLISGTGALGSRSLFSP 53
DB 215 GQFGNDANFLTFLEPRASAPIVGTWAPTQCSPRDTAA-----GPLNAGAGPLATIQI--L 267
QY 54 RNSNADSVSDRIPLPTNPSR--LAAATSETCLGGFVLHD-----KGPLDILNTQIGP 107
DB 268 RQAVA-TVSYLDGGSVNTNGGFEFTINATVPTPSGQVQFTRDGEDVGAPVDLVN--GK 323
QY 108 SAFRVEVQADGTHAAIGEKNGLEV-----SVTLSPQEWSSLSQSIDTEGKNRFVFTGG 159
DB 324 ASLTQSLDLDGDYAVEAKFLGAEFFNPSSAAKTVTVTSQDIQTTSVT----- 371
QY 160 RGGSGHPMVTVASDIAEARTRILAKLDDNNGGGRQPKDVTDRSVGSGASGIDGVVSET 219
DB 372 --GPDH-----DAYRDQPNLTKRKEPVGSGVFAFEVD--GTPVGTADVMDGAAVLP 421
QY 220 HTSTTNSVSRSDPKFVWSGAIAGLAGLAATGIAQALAL-----TPRPDDPTTDP 271
DB 422 HTFTTNGTHR-----VIARYSG--REGISPSVSLQYPVSVTEAPADVATTITV 468

```

; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268,347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 36
; LENGTH: 2411
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-268-347-36

Query Match          5.0%; Score 139.5; DB 4; Length 2411;
Best Local Similarity 21.5%; Pred. No. 0.0096;
Matches 125; Conservative 83; Mismatches 228; Indels 145; Gaps 28;

QY      6  LGNNVGNH-----LIPPAPLPSQTDGAARG---GTGHLISSIT--GALGSRSLFSPLRNS 56
DB      1161  LGGTANGRNDGTGVINKOGLTITLANGAAGAGTDASNGNTISVTKOGISAGNKBITVVKSA 1220
QY      57  MADSDVSRDIPGLPTNPRLAAATSTCLLGGFEVLHDKGPLDILNTQIGPSAFRVEVQA 116
DB      1221  LKTYKDTQNTAG-ATQP---AANTA-EVA-----KQDLVDLTKPATGAAGNGADAKA 1267
QY      117  -DGTHAAIGEKNGL-----EVSVTLSPQEW---SSLQSIDTEGKNRFVFTGGRGGSG 164
DB      1268  PDITTAATVGDLRGLGWLVSAAKKTADETQKEFAHFAVKNAEVEFVGKNGATVSAKTDNNG 1327
QY      165  HPWTVVASDIABAR-----TRILAKLDPDNHGGRQPKVDVTRSVGVGSAGSIDDGVV 216
DB      1328  KHTVTIT--DVAEAKVGDGLEKDTDGKIKLKVNDNTDGNLLTVDA--TKGASVAKGEFNAV 1384
QY      217  SETHTST-TNSSVR-----SDPKFVSVGAIAGLAGLAATGIAQALALT 260
DB      1385  TDATTAGGTANERGKVVKSGNGATATETDKKVAIVGDVAKAIND-AATFV-----K 1437
QY      261  PEPDDPTTTDPQANAAESATK--DQLTQEAFFKNPENOKVNID-----ANGNAIPS 310
DB      1438  VENDDSATIDDSPTDGDANDALKAGDTLTLKAGK---NLKVKEDGKNIITFALANDLSVKS 1494
QY      311  GELXDDIVGIAQQAKEAGEVARQQAESNAQAQRYEDQHARRQBELQLSSGIGYGLSS 370
DB      1495  ATVSDKL-----SLGTNGNKVNITSDTKGLNFAKDSKTGDDANIHL-NGIASTLTD 1544
QY      371  ALIVAGGIGAGVTTALHRRNPABEQTTTTTHTVVQQQTGGIIPQHKVALMPQERRRFSDR 430
DB      1545  TL-----LNSGATTNL-----GG-----NGITDNKKKAASV 1571
QY      431  RDSQGSVASTHSDSSSESVNPPYAEVGGARNSLSAHQPBEEHIYDEVAADPGYSVIONFSG 490
DB      1572  KD-----VLNAGWNVGVKPA-SANNOVENI-DFVAT---YDVTDFVSG 1610
QY      491  SGPVTCRLTGTPGGQTGSTYALLANSGLRLNGMGLTSGGE 531
DB      1611  DKDTSVTVESKONGKRTEVKIGAKTSVIKDHNGKLTGKE 1651

RESULT 6
US-09-377-155-33
; Sequence 33, Application US/09377155
; Patent No. 6197312
; GENERAL INFORMATION:
; APPLICANT: PEAK, Ian Richard Anselm
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/377,155
; CURRENT FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2

```

;; PRIOR FILING DATE: 1997-12-12
;; NUMBER OF SEQ ID NOS: 33
;; SOFTWARE: Patentin Ver. 2.0
;; SEQ ID NO 33
;; LENGTH: 2353
;; TYPE: PRT
;; ORGANISM: Haemophilus influenzae
US-09-377-155-33

Query Match 4.9%; Score 138.5; DB 3; Length 2353;
Best Local Similarity 22.0%; Pred. No. 0.011;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;
QY 6 LGNNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
DB 1163 LGGTANG-----RNDTGVINKD-----LTITLANGAAAGTDSN 1198
QY 66 IPGLPTNPRLAAATSETCLLGGPEVLHDKGPL-----DILNTQIGPSAFRVEVQADGTHA 121
DB 1199 -----GNTISVTKDGISAGNKEITNVKSALTKYKDTQNT-----ADETDQKEPHA 1243
QY 122 AIGKXNGLEVSVTLSPQEWSSLSQIDTEGKNRFVFTGGRGGSGHPMTVVASDIAEAR--- 178
DB 1244 AV--KNANEVEFV-----GKNGATVSAKTDNNNGKHTVTI--DVAEAKVGD 1284
QY 179 -----TRILAKLDPNHHGGRQPKVDTRSVGVSGSAGIDGCVVSETHST-TNSSVR-- 229
DB 1285 GLEKDTDGKIKLVNDTGNLLTVDA--TKGASVAKGFEFNAVTTDTAATGQTNANERGV 1343
QY 230 -----SDPKFWVSUGAIAAGLAGLAATGIAQALALTPEDPTTTDPQAAANA 277
DB 1344 VVKGSGNATATETDKKVVATVGVAKAIND-AATFV-----KVENDDSATIDDSPTDDG 1396
QY 278 AASATK--DQLTOEAFKNPENOKNID-----ANGNAIPSGELXDDIVEQIAQAQAKE 327
DB 1397 ANDALKAGDTLTKAGK--NLKVRDGNITFALANDLSVKSATVSDKL-----S 1444
QY 328 AGEVARQAAVESNAQAQOQRYEDQHARRQEEELQSSGIGYGLSSALIVAGGIGAGVTTALH 387
DB 1445 LGTNGKNVNTSDTKGLNFAKSKTGDDANIHL-NGIASTLTDTL-----LNSGATTNL- 1497
QY 388 RRNPQAEQTTTTTHTVVQOQTGGIPQHKVAMPQERRFRDRDRSQGSVASTHWSDS 447
DB 1498 -----GG-----NGITDNEKKRAASVKD----- 1515
QY 448 EVVNPYAEVGGARNLSAHOPEEHIYDEVAADPGYSVIQNFSGSGPVTRGLIGTPCGQIQ 507
DB 1516 -VLNAGMNVRGVKA-SANNQVENI-DFVAT---YDVFVSGDKDXTTSVTVESKDNKGR 1569
QY 508 STYALLANSGLRLGMMGLTSGGE 531
DB 1570 TEVKIGAKTSVIKDHNGKLTGKE 1593

RESULT 7

US-08-913-942-4
; Sequence 4, Application US/08913942
; Patent No. 6200578
; GENERAL INFORMATION:
; APPLICANT: St. Geme, Joseph
; APPLICANT: Barenkamp, Stephen J.
; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible

;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/913,942
;; FILING DATE: 29-DEC-1997
;; CLASSIFICATION: 514
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/409,995
;; FILING DATE: 24-MAR-1995
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US96/4031
;; FILING DATE: 22-MAR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Vance, Dolly A.
;; REGISTRATION NUMBER: 39,054
;; REFERENCE/DOCKET NUMBER: A-61053-1/RPT/RMS/DAV
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 781-1989
;; TELEFAX: (415) 398-3249
;; TELEX: 910 277299
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2353 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-913-942-4

Query Match 4.9%; Score 138.5; DB 3; Length 2353;
Best Local Similarity 22.0%; Pred. No. 0.011;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;
QY 6 LGNNVNGNHLIPAPPPLPSQTDGAARGGTGHLISSTGALGSRSLFSPLRNSMADSVDSRD 65
DB 1163 LGGTANG-----RNDTGVINKD-----LTITLANGAAAGTDSN 1198
QY 66 IPGLPTNPRLAAATSETCLLGGPEVLHDKGPL-----DILNTQIGPSAFRVEVQADGTHA 121
DB 1199 -----GNTISVTKDGISAGNKEITNVKSALTKYKDTQNT-----ADETDQKEPHA 1243
QY 122 AIGKXNGLEVSVTLSPQEWSSLSQIDTEGKNRFVFTGGRGGSGHPMTVVASDIAEAR--- 178
DB 1244 AV--KNANEVEFV-----GKNGATVSAKTDNNNGKHTVTI--DVAEAKVGD 1284
QY 179 -----TRILAKLDPNHHGGRQPKVDTRSVGVSGSAGIDGCVVSETHST-TNSSVR-- 229
DB 1285 GLEKDTDGKIKLVNDTGNLLTVDA--TKGASVAKGFEFNAVTTDTAATGQTNANERGV 1343
QY 230 -----SDPKFWVSUGAIAAGLAGLAATGIAQALALTPEDPTTTDPQAAANA 277
DB 1344 VVKGSGNATATETDKKVVATVGVAKAIND-AATFV-----KVENDDSATIDDSPTDDG 1396
QY 278 AASATK--DQLTOEAFKNPENOKNID-----ANGNAIPSGELXDDIVEQIAQAQAKE 327
DB 1397 ANDALKAGDTLTKAGK--NLKVRDGNITFALANDLSVKSATVSDKL-----S 1444
QY 328 AGEVARQAAVESNAQAQOQRYEDQHARRQEEELQSSGIGYGLSSALIVAGGIGAGVTTALH 387
DB 1445 LGTNGKNVNTSDTKGLNFAKSKTGDDANIHL-NGIASTLTDTL-----LNSGATTNL- 1497
QY 388 RRNPQAEQTTTTTHTVVQOQTGGIPQHKVAMPQERRFRDRDRSQGSVASTHWSDS 447
DB 1498 -----GG-----NGITDNEKKRAASVKD----- 1515
QY 448 EVVNPYAEVGGARNLSAHOPEEHIYDEVAADPGYSVIQNFSGSGPVTRGLIGTPCGQIQ 507
DB 1516 -VLNAGMNVRGVKA-SANNQVENI-DFVAT---YDVFVSGDKDXTTSVTVESKDNKGR 1569
QY 508 STYALLANSGLRLGMMGLTSGGE 531
DB 1570 TEVKIGAKTSVIKDHNGKLTGKE 1593

RESULT 8

US-09-669-974-33
; Sequence 33, Application US/09669974
; Patent No. 6333173
; GENERAL INFORMATION:
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: JENNINGS, Michael Paul
; APPLICANT: MOXON, E. Richard
; TITLE OF INVENTION: NOVEL SURFACE ANTIGEN
; FILE REFERENCE: 065064/0128
; CURRENT APPLICATION NUMBER: US/09/669,974
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: US 09/377,155
; PRIOR FILING DATE: 1999-08-19
; PRIOR APPLICATION NUMBER: PCT/AU98/01031
; PRIOR FILING DATE: 1998-12-14
; PRIOR APPLICATION NUMBER: GB 9726398.2
; PRIOR FILING DATE: 1997-12-12
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patent in Ver. 2.0
; SEQ ID NO 33
; LENGTH: 2353
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
US-09-669-974-33

Query Match 4.9%; Score 138.5; DB 4; Length 2353;
Best Local Similarity 22.0%; Pred. No. 0.011;
Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;
QY 6 LGNNVGNHILPPAPPPLPSQTDGAARGGTGHLISSTGALGRSLPSLRNSWDSVDRD 65
DB 1163 LGGTANG-----RNDTGTINKDG-----LTITLANGAAAGTASN 1198
QY 66 IPGLTNPRLAAATSETCLGGFVLDKGPL-----DILANTQGPSAFRVEVQADGTHA 121
DB 1199 -----GNTISVTKDGISAGNKEITNVKSALTKYDQTNT-----ADETQDKPHA 1243
QY 122 AIGKNGLEVSVTLSPQEWSSLSQSDTEGKRFVFTGGRGSGHGMVTVASDIABAR--- 178
DB 1244 AV--KNAVEFEV-----GKNGATVSAKTDNNGKHTVTL--DVAEAKVGD 1284
QY 179 -----TRILAKLPDNNHGRQPKVDTRSVGVSAGSIDDGVVSETHST--TNSSVR--- 229
DB 1285 GLEKDTGKIKLVKVDNTGNNLLTVDA--TKGASVAKGEFNAVTTDTATTAQGINANERGV 1343
QY 230 -----SDPKFVSVCAIAAGLAGLAATGIAQALALTPPPDPTTIDPQAAANA 277
DB 1344 VVKGNGATATETDKKVVATVGDVAKAIND--AATFV-----KVENDDSATIDDSPTDDG 1396
QY 278 AESATK--DQLTOEAFKPNOKVNID-----ANGNAIPSGELXDDIVEQIAQAKE 327
DB 1397 ANDALKAGDTLLKAGK---NLKVRDGNKITFALANDLSVKSATVSKL-----S 1444
QY 328 AGEVARQAVESNAQAQRYEDQHARRQBEELQSSGIGYGLSSALIVAGGIGAGVTTALH 387
DB 1445 LGTNGKNVITSDTKGLNFAKDSKTGDDANIHL--NGIASLTLDL-----LNSGATTNL- 1497
QY 388 RNQPAEQTTTTTTHVVVQQQTGGIPQHKVAMPQERRRRFRDRRSQGSVASTHWSDS 447
DB 1498 -----GG-----NGITDNEKRAASVKD----- 1515
QY 448 EYVNPVAVGGARNSLSAQHPEEHLVDEVAADPGYSVIONFSGSGSPVTRGLIGTPGGIQ 507
DB 1516 -VLNAGWNRGVKPA--SANNQVENI--DFVAT---YDVFVSGDKDTSVTVESKDNKR 1569
QY 508 STYALLANSGLRLGMGGITSGGE 531
DB 1570 TEVKIGAKTSVIKDHNGKLPFGKE 1593

RESULT 9

US-09-252-991A-30227

; Sequence 30227, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30227
; LENGTH: 2736
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30227

Query Match 4.9%; Score 137; DB 4; Length 2736;
Best Local Similarity 21.5%; Pred. No. 0.019;
Matches 135; Conservative 56; Mismatches 275; Indels 162; Gaps 28;
QY 2 PIGNLNNVGNHILPPAPPPLPSQT-----DGAARGGTGHLISSTGALGRSLPSPLR 54
DB 477 PIGQTSADANGNWSFTPGSQLPDGTVVNVVWARDAGNSSPATSTVDCVAPNAPVVEPSN 536
QY 55 NS-----MADSVDSRDIPGLPTNPRLAA-----ATSETCLGGEVLHDKGPLDILN 102
DB 537 GSELGTAEPGSSVTLTDGNGNPIGQTTADANGNWSFTPTPLPDGTVV--NVVARDAG 594
QY 103 TQIGPSAFRVEVQADGTHAAIGKNGLEVS-----VTLSPQEWSSLSQIDTEGKRF 154
DB 595 NSSPASPVTVDAPAT--PTVDPNGTTLSTGABGATVTLTDGNGNPIGQVTADGSGNW 653
QY 155 VFTGG-----RGGSGHPMTVVASDIABARTRILAKLPDND---HGRQ 194
DB 654 TFTPSTPLPNTVVNATATDPSGNASSPASVTVDPVAPATPVW---NPSNGTTLSGTAE 709
QY 195 PKVDVTRSVGVSAG--IDGVVSETHSTT---NSSVRSDPKFVSVGAIAGLAGLA 249
DB 710 PGATVTLADGNGNPIGQVTADGSGNWSFTPTPLPNTGTV-----VNATATDASGNT 760
QY 250 ATG---IAQALALTP--EPDDPTTDDQANAASATKDQLTOEAFKPNOKVNIDA 303
DB 761 SAGSVTVDSVAPATPVINPNSGTT-----LSGTABPGSSVTLT--DGNGNPIGO--VTADG 813
QY 304 NGN--AIPSGELXDDIVEQIAQAKEAGEVARQAVESNAQAQRYEDQHARRQBEELQLS 361
DB 814 SGNWSFTPTPLADGTVVN--ATATDPAGNTSGQSGTT-----VDGVAPTTPTVNL 863
QY 362 SGI-----GYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTTHVV 405
DB 864 NGSSLSGTABPGSTVILLTDGNGNPIAEVTADGSGNWTVT-----PSTPIANGTVNVV 916
QY 406 QQQTGGIPQHKVAMPQERRRRFRDRRSQGSVASTHWSDSSEVVNPVAVGGARNSLSA 465
DB 917 AQDAAG-----NSSPGASVTVDSQAPAPVNP-----SNGTTLUG 952
QY 466 HQPEEHLVDEVAADPGYSVIONFSGSGPV-----TGRLLGTGCGQIQSTYALLANS 518
DB 953 -----TABPGATVSLTDGNGNPIGQVTADGSGNWSFTPTPLANGTVNVNATASD 1001
QY 519 LRLGMGGITSGGETAVSSV--NAAPTGP 545
DB 1002 ---PTGNTSAPASTTVDSVAPAPVNP 1026

RESULT 10

US-09-328-352-4412
; Sequence 4412, Application US/09328352
; Patent No. 6562958

Db 1569 TEVKIGAKTSVIKHNGKLTGKE 1592

RESULT 12

US-08-685-467-4

; Sequence 4, Application US/08685467

; Patent No. 6060059

; GENERAL INFORMATION:

; APPLICANT: St. Geme III, Joseph W.

; APPLICANT: Barenkamp, Stephen J.

; TITLE OF INVENTION: HAEMOPHILUS ADHESION PROTEINS

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESS: Flehr, Hobbach, Test, Albritton & Herbert

; STREET: Four Embarcadero Center, Suite 3400

; CITY: San Francisco

; STATE: California

; COUNTRY: United States

; ZIP: 94111-4187

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/685,467

; FILING DATE: 22-JUL-1996

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/409,995

; FILING DATE: 24-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Silva, Robin M.

; REGISTRATION NUMBER: 38,304

; REFERENCE/DOCKET NUMBER: A-61053-2/RPT/RMS

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 781-1989

; TELEFAX: (415) 398-3249

; TELEX: 910 277299

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1912 amino acids

; TYPE: amino acid

; STRANDEDNESS: unknown

; TOPOLOGY: unknown

; MOLECULE TYPE: protein

US-08-685-467-4

Query Match 4.8%; Score 134.5; DB 3; Length 1912;

Best Local Similarity 22.0%; Pred. No. 0.016;

Matches 124; Conservative 62; Mismatches 207; Indels 171; Gaps 26;

QY 6 LGNVNNGHILPPAPLPPLSQTDCGARGGTGHLISSTGALGSRSLPSPLNSMADSVDSRD 65

Db 1162 LGGTANG-----RNDTGTINKDG-----LTLTLANGAAGTDSAN 1197

QY 66 IGLPTNPSRLAAATSETCLGGFVLHDKGPL-----DILNTQIGPSAFRVEVQADGTHA 121

Db 1198 -----GNTISVTKGDISAGNKEITNVKSALTKYKDTQNT-----ADETDKEPHA 1242

QY 122 AIGKNGLEVSVTLSPQEWSSLSQSDTEGKRFVFTGGRGSGHGMVTVASDIABAR--- 178

Db 1243 AV--KNANEVEFV-----GKNGATVSAKTDNNGKHTVI--DVAEAKVGD 1283

QY 179 -----TRILAKLPDPNHGQPKVDVTRSVGSGSAGIDGCVSVSTHST-TNSSVR--- 229

Db 1284 GLEKTDGKIKLVNTDGNLLTVDA-TKGASVAKGEFNAVTTTATTAAQGTNANERGV 1342

QY 230 -----SDPKFVSVSCAIAAGLAGLAATGIAQALALTPEPDPTTDPDQANA 277

Db 1343 VVKNGSGATATETDKKATVGDVAKAIND-AATFV-----KVENDDSATIDDSPTDDG 1395

QY 278 ABSATK--DQLTQEAFFKNPENQKNVID-----ANGNAIPSGELXDDIVEQIAQAKE 327

Db 1396 ANDALKAXDILTLRAGK---NLKVRDGRKNTTFALANDLSVKSATVSDKL-----S 1443

QY 328 AGEVARQOAVESNAQAQRYEDQHARRQEELQSSGIGYGLSSALIIVAGGIGAGVTTALH 387

Db 1444 LGTNGKNVITSDTKGLNFADKSDTGDANIHL-NGLASTLTDTL-----LNSGATNLT- 1496

QY 388 RRNPQAEQTTHHTVVVQQTGGIPOHKVALMPQERRRFRDRDSQGSVASTHWSDESS 447

Db 1497 -----GG-----NGITDNEKKXAA-----SVK 1513

QY 448 EVVNPYAEVGCARNLSHQPEEHYDEVAADPCSVIQNESGSGPVTGRLIGTPGQIG 507

Db 1514 DVLNAGMNVRGVKA-SANNQVENI-DFVAT---YDTVDFVSGDKDTTSTVTVESKDNKR 1569

QY 508 STYALLANSGLRLGMGLTSGGE 531

Db 1569 TEVKIGAKTSVIKHNGKLTGKE 1592

RESULT 13

US-09-252-991A-21021

; Sequence 21021, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 21021

; LENGTH: 375

; TYPE: PRT

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-21021

Query Match 4.8%; Score 134; DB 4; Length 375;

Best Local Similarity 23.8%; Pred. No. 0.0016;

Matches 81; Conservative 44; Mismatches 158; Indels 58; Gaps 13;

QY 195 PKDVTTRSVGVGSASGIDGTVVSETHSTTSSVRSR-D-PKFWVSVGATAAGLAGLAATGI 253

Db 6 PEAFQVVELGIGQGEAGAQLVPKDRHQGSADDDADHPVHLALHR-AVHPAGLATAGE 64

QY 254 AQAALTPEPDDPTTTDPDQAANAASATKDQTOEAFKPNPENQKNVIDANGNAIP---- 309

Db 65 DDADAENVAADRTRAHPEDLRRDAQ-ADREGTDAAGRHAEQ-----DAEGGAGPAGGP 118

QY 310 -----SGELXDDIVEQIA-QQAKEAGEVARQQAQRYEDQHARRQEELQ-LSS 362

Db 119 EVPCGTGEADLRALDQAEQQAQAGHAQHRYLEAPG-GEQGEROHRQAQAGIQAQAS 177

QY 363 GIGVGLSALIIVAGGIGAGVTTALHRRNQPAEQTTTTTHTVVQ--QOTGGIPQHKVALM 420

Db 178 QQAFQ-----AQCGAGRGADPQCAQERQQADLHADPGAADGGEHGFVQGKHRTD 229

QY 421 PQERRRSDRDSQGSV-----ASTHMSDSSEVVNPYAEVGA----- 459

Db 230 GEERQVAANHQARAEAGDDTDATQLOATFAGPAQADVAGPEGAAQGAEBEADDDHSHV 289

QY 460 ---RNSLSAHQPEEHYDEVAADPGYSVIQ---NFGSGP 493

Db 290 DPVQRALKADQPEVAPVILVLAEPGQHVAAPEFAGHRP 330

RESULT 14

US-08-185-432-19

[illegible]

Search completed: December 16, 2003, 08:57:53
Job time : 28.7653 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 16, 2003, 08:50:49 ; Search time 53.0171 Seconds
(without alignments)
1643.636 Million cell updates/sec

Title: US-09-189-415A-2
Perfect score: 2800
Sequence: 1 MPIGNLGNVNGNHLIPPAP.....GETAVSSVNAAPTGPVRV 549

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description
1	2798	99.9	549 20	AA1980
2	1492	53.3	559 20	AA1981
3	530	18.9	107 21	AA1982
4	165	5.9	30 20	AA1983
5	151	5.4	934 22	AA1984
6	149	5.3	831 19	AA1985
7	145.5	5.2	848 22	AA1986
8	145.5	5.2	2586 22	AA1987
9	144	5.1	883 22	AA1988

10	143.5	5.1	1046	22	ABB67028	Drosophila melanog
11	141.5	5.1	1488	22	AAG91741	C glutamicum prote
12	141	5.0	580	18	AAW10660	Endoglycoceramidas
13	140.5	5.0	1472	22	ABBS9423	Drosophila melanog
14	139.5	5.0	985	23	ABP43746	2 PUMH2 protein.
15	139.5	5.0	2411	21	AA193860	Haemophilus influe
16	138.5	4.9	2353	17	AAR99393	Haemophilus adhesi
17	138	4.9	1455	22	ABBS3040	Drosophila melanog
18	138	4.9	2857	22	ABBS4188	Drosophila melanog
19	136.5	4.9	3096	22	AA146770	FANCI3 protein fr
20	136.5	4.9	4630	18	AAW19629	Streptomyces venez
21	136	4.9	1064	22	ABBS9068	Drosophila melanog
22	136	4.9	1569	22	AAG98842	E. coli growth and
23	136	4.9	2086	22	AAU34143	Staphylococcus aur
24	136	4.9	5795	22	AAU37017	Staphylococcus aur
25	135.5	4.8	1138	23	ABJ05434	Human breast cance
26	135	4.8	1045	22	ABG08332	Novel human diagno
27	135	4.8	1532	21	AB40945	Human ORFX ORF709
28	135	4.8	1905	22	ABBS4925	Drosophila melanog
29	135	4.8	2344	22	AAU37120	Staphylococcus aur
30	134	4.8	828	22	ABBS68368	Drosophila melanog
31	133.5	4.8	620	22	ABBS6938	Drosophila melanog
32	133.5	4.8	818	22	ABBS70219	Drosophila melanog
33	133.5	4.8	1978	20	AAU27230	Amino acid sequenc
34	133	4.8	745	22	AAU25461	Human mddt protein
35	133	4.8	1596	22	ABW70845	Drosophila melanog
36	132.5	4.7	1026	18	AAW37490	Caulobacter cresce
37	132.5	4.7	1026	21	AAU44757	Caulobacter cresce
38	132.5	4.7	1026	24	AAE34374	Caulobacter cresce
39	132.5	4.7	1126	22	ABBS0460	Drosophila melanog
40	132.5	4.7	4630	21	AAU77177	S. venezuelae vep
41	131.5	4.7	707	24	ABJ26513	Aspergillus fumiga
42	131.5	4.7	1026	15	AA148993	rsaA S-lyase prote
43	130.5	4.7	950	22	ABBS2035	Drosophila melanog
44	130.5	4.7	1026	17	AA194014	Caulobacter S-layer
45	130.5	4.7	2304	22	ABBS0219	Drosophila melanog

ALIGNMENTS

RESULT 1
AA1980220
ID AA1980220 standard; Protein; 549 AA.
AC AA1980220;
XX
DT 16-AUG-1999 (first entry)
DE EPEC E. coli translocated intimin receptor (Tir).
KW Tir; translocated intimin receptor; Hp90; enteropathogenic;
XX EPEC; infection; diagnosis; vaccine.
OS Escherichia coli.
FH Key Location/Qualifiers
FT Domain 234..253
FT /note= "putative transmembrane domain"
FT Domain 364..386
FT /note= "putative transmembrane domain"
FT Misc-difference 180
FT /note= "encoded by AAA"
FT Misc-difference 314
FT /note= "given as Xaa in the specification; Lys
is deduced from the DNA sequence"
XX WO9924576-A1.
XX 20-MAY-1999.
XX 10-NOV-1998; 98WO-CA01042.
XX

```
PR 12-NOV-1997; 97US-0065130.
XX (UYBR-) UNIV BRITISH COLUMBIA.
XX Devinney R, Finlay BB, Kenny B, Stein M;
DR WPI; 1999-337712/28.
DR N-PSDB; AAX58858.
XX
XX New translocated intimin receptor useful for treating infection by
XX enteropathogenic or enterohemorrhagic Escherichia coli
XX
XX Claim 6; Page 55-58; 91pp; English.
XX
XX The present sequence represents Tir, a novel translocated intimin
XX receptor (formerly termed Hp90) from an enteropathogenic
XX Escherichia coli (EPEC) strain. The sequence was deduced from an
XX isolated tir polynucleotide (see AAX58858). Tir proteins are
XX secreted by attaching and effacing pathogens such as EPEC and EHEC
XX (see AAY06221) E. coli. The bacterial pathogens insert their own
XX receptors into mammalian cell surfaces, to which the pathogen then
XX adheres to trigger additional host signaling events and actin
XX nucleation. Diagnosis of disease caused by pathogenic E. coli can
XX be performed by use of antibodies that bind to Tir to detect the
XX protein or the use of nucleic acid probes for detection of nucleic
XX acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
XX peptides, a recombinant method for producing recombinant Tir,
XX antibodies which bind to Tir, and a kit for the detection of
XX Tir-producing E. coli are provided. A method of immunising a host
XX with Tir to induce a protective immune response is also provided.
XX In addition, Tir fusion proteins can be used in attenuated E. coli
XX to induce a cell-mediated immune response to other polypeptides,
XX e.g. antigens. A method for screening for compounds which
XX interfere with the binding of bacterial pathogens to their
XX receptors is further provided.
XX
XX Sequence 549 AA;
XX
XX Query Match 99.9%; Score 2798; DB 20; Length 549;
XX Best Local Similarity 99.8%; Pred. No. 2.6e-210;
XX Matches 548; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 1 MPIGNLNNVNGNHLIPAPPPLPSQTDGAARGCTGHLISSTGALGSRSLFSPLRNSMADS 60
DB 1 MPIGNLNNVNGNHLIPAPPPLPSQTDGAARGCTGHLISSTGALGSRSLFSPLRNSMADS 60
QY 61 VDSRDI PGLTPNPSRLAATSETCLLGGFEVLHDKGPDILNTOIGPSAFRVEVQADGTH 120
DB 61 VDSRDI PGLTPNPSRLAATSETCLLGGFEVLHDKGPDILNTOIGPSAFRVEVQADGTH 120
QY 121 AAIKGNLGVSVTLSPQEWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEARTR 180
DB 121 AAIKGNLGVSVTLSPQEWSSLSQSDTEGKRFVFTGGRGSGHPMTVVASDIAEARTR 180
QY 181 ILAKLDPDNHGRQPKVDTRSVGVSGASGIDDGWVSEHTTSTNSSVRSDPKFWVSGA 240
DB 181 ILAKLDPDNHGRQPKVDTRSVGVSGASGIDDGWVSEHTTSTNSSVRSDPKFWVSGA 240
QY 241 IAAGLAGLAATGIAQALATPEPDDPTTDPQANAASATKDQLTQEAFAKNPENQKN 300
DB 241 IAAGLAGLAATGIAQALATPEPDDPTTDPQANAASATKDQLTQEAFAKNPENQKN 300
QY 301 IDANGNAIPSGELXDDIVQIAQAKEAGEVARQOAVESNAQOQRYEDQHARRQELQL 360
DB 301 IDANGNAIPSGELKDDIVQIAQAKEAGEVARQOAVESNAQOQRYEDQHARRQELQL 360
QY 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHTVVQQQTGGIPQHKVALM 420
DB 361 SSGIGYGLSSALIVAGGIGAGVTTALHRRNQPAEQTTTTTHTVVQQQTGGIPQHKVALM 420
QY 421 POERRRFRDRSQSVASTHWSDSSESVNYPYAEVGGARNLSAHOPEEHLYDEVAADP 480
DB 421 POERRRFRDRSQSVASTHWSDSSESVNYPYAEVGGARNLSAHOPEEHLYDEVAADP 480
```

```
QY 481 GYSVIQNFSGSPVTRGLTGPQGIQSTYALLANSGLRLGMGLTSGGETAVSSVNA 540
DB 481 GYSVIQNFSGSPVTRGLTGPQGIQSTYALLANSGLRLGMGLTSGGETAVSSVNA 540
QY 541 PTPGPVRPV 549
DB 541 PTPGPVRPV 549
XX
XX RESULT 2
XX ID AAY06221 standard; Protein; 559 AA.
XX AC AAY06221;
XX XX 16-AUG-1999 (first entry)
XX DE EHEC E. coli translocated intimin receptor (Tir).
XX KW Tir; translocated intimin receptor; Hp90; enterohaemorrhagic;
XX EHEC; infection; diagnosis; vaccine.
XX OS Escherichia coli.
XX FH Key Location/Qualifiers
XX FT Misc-difference 453 /note= "encoded by codon of 1 apparent nucleotide,
XX FT causing frameshift in the DNA sequence"
XX FT
XX FN WO9924576-A1.
XX PD 20-MAY-1999.
XX PF 10-NOV-1998; 98WO-CA01042.
XX PR 12-NOV-1997; 97US-0065130.
XX XX (UYBR-) UNIV BRITISH COLUMBIA.
XX FI Devinney R, Finlay BB, Kenny B, Stein M;
XX DR WPI; 1999-337712/28.
XX N-PSDB; AAX58859.
XX PT New translocated intimin receptor useful for treating infection by
XX enteropathogenic or enterohemorrhagic Escherichia coli
XX PS Claim 7; Page 55-58; 91pp; English.
XX
XX The present sequence represents Tir, a novel translocated intimin
XX receptor (formerly termed Hp90) from an enterohaemorrhagic
XX Escherichia coli (EHEC) strain. The sequence was deduced from an
XX isolated tir polynucleotide (see AAX58859). Tir proteins are
XX secreted by attaching and effacing pathogens such as EHEC and EPEC
XX (see AAY06220) E. coli. The bacterial pathogens insert their own
XX receptors into mammalian cell surfaces, to which the pathogen then
XX adheres to trigger additional host signaling events and actin
XX nucleation. Diagnosis of disease caused by pathogenic E. coli can
XX be performed by use of antibodies that bind to Tir to detect the
XX protein or the use of nucleic acid probes for detection of nucleic
XX acids encoding Tir polypeptide. Isolated Tir nucleic acids, Tir
XX peptides, a recombinant method for producing recombinant Tir,
XX antibodies which bind to Tir, and a kit for the detection of
XX Tir-producing E. coli are provided. A method of immunising a host
XX with Tir to induce a protective immune response is also provided.
XX In addition, Tir fusion proteins can be used in attenuated E. coli
XX to induce a cell-mediated immune response to other polypeptides,
XX e.g. antigens. A method for screening for compounds which
XX interfere with the binding of bacterial pathogens to their
XX receptors is further provided.
XX
XX Sequence 559 AA;
```

```
Query Match      53.3%; Score 1492; DB 20; Length 559;
Best Local Similarity 54.7%; Pred. No. 4.2e-108;
Matches 322; Conservative 64; Mismatches 133; Indels 70; Gaps 12;

QY 1 MPIGNLGNVNGNHLIPAPPLPSOTDGAARGGTGHLISSTGALGSRSLFPLRNSMADS 60
DB 1 MPIGNLGNVNGNHLIPAPPLPSOTDGA--GGRQLINSTGPLGSRALFTFVRNSMADS 58

QY 61 VDSR--DIPGLPTNPSRLAAATSETCLGGFEVLHDKGLDILNTOIGPSAFRVEQADG 118
DB 59 GDNRASDPGLPVPNMLAA--SETLNDGFEVLHDKGLDILNTOIGPSAFRVEQADG 116

QY 119 THAIGEKNGLVSVTLSPQWSSLSQIDTEGKRFVFTGGRGSGHGMVTVASDIABAR 178
DB 117 KHIAVGQRNGVETSVLSQDEYARLSIDPEGKQFVFTGGRGSGHGMVTVASDIABAR 176

QY 179 TRIIAKLDPDNGHGRQKDVTRSVGVSAGI-----DGV--VSTHTSTNNSVRS 230
DB 177 QRILLELEPKGTG-----ESKGAGESKGVGELRESNSGAENTTTTQSTSTSLRS 227

QY 231 DPKFWSVGALAAAGLAGLAATGIAQALATPEPDPTTTDPDAANAASATKDQLTQEA 290
DB 228 DPKLWLAGVTATGLIGLAATGIVQALATPEPDPTTTDPDAASAATETATRDQLTKEA 287

QY 291 FKNPENQKVNIDANGNAIPSGELXDDIVEQIAQAQAGEVARQAVESNAQAQORYEDQ 350
DB 288 FQNPDNQKVNIDELGNAPSGVLKDDVVAIEEQAAGEAKQAIENNAQAQKRYDEQ 347

QY 351 HARQEELQLSGIGYGLSSALIIVAGTGAGVTTHLHRRNPAPQOTTTTT-----HT 403
DB 348 QAKQEELKVSAGYGLSGALILGGGIGVAVTAALHRRKNQPVQETTTTTTTTTTSART 407

QY 404 VVQQTGIGIQHVKALMPQERRRSDRRDSQGSVASTHWSDSSEVNPNYAEVGG--AR 460
DB 408 VENPANNTPAQGNVDTFGSEDTWESRRSSMASTSTFTDSS-----IGGFCRIR 458

QY 461 NSLSAHQ-----PEEHYDVAADPGVSVIQTGSGGPVTRGLIG 500
DB 459 MMLKHCRCMIRCRILLIRLFRWIGIQISVVYSTIQHPP-----RDTDNG---ARLLG 510

QY 501 TFGQGIQSTYALLANSGLRLGMLGSLTGGTAVSSVNAAPTPGPRFV 549
DB 511 NPSAGIQSTYARLALSGLRHDMLGSLTGGNSAVNTSNPPAPGSHRFV 559

RESULT 3
AAB20576
XX AAB20576 standard; protein; 107 AA.
XX
XX AAB20576;
XX
DT 08-DEC-2000 (first entry)
XX
DE Intimin C-terminal Tir binding domain amino acid sequence.
XX
KW Intimin; Tir binding domain; inhibitor; intimin adhesion; screening;
KW Tir-independent eukaryotic cell binding activity; bacterial infection;
KW diarrhoea; antibacterial.
XX
OS Unidentified.
XX
PN WO200045173-A1.
XX
PD 03-AUG-2000.
XX
PF 31-JAN-2000; 2000WO-GB00254.
XX
PR 29-JAN-1999; 99GB-0001897.
XX
PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.
XX
PI Frankel GM, Matthews SJ, Hale CB, Dougan G;

WPI; 2000-499357/44.
Screening for inhibitors of intimin binding to eukaryotic cells, for
use in diagnosing, preventing and treating bacterial infections,
especially Escherichia coli O157:H7.
Claim 8; Page 76; 96pp; English.

The present invention describes a method of screening for an inhibitor
of intimin binding to eukaryotic cells. The method comprises exposing an
intimin polypeptide having a Tir-independent cell binding activity to
test agents, and obtaining an inhibitor based on its ability to bind the
polypeptide. The inhibitors are used in the prevention, treatment and/or
diagnosis of bacterial infections, preferably by enteropathic and/or
enterohaemorrhagic Escherichia coli, Shiga toxinogenic E. coli, Hafnia
alvei or Citrobacter freundii, or especially E. coli O157:H7. The
infections cause a histopathological effect known as attachment and
effacement on intestinal epithelial cells. The inhibitors can be used
to produce food supplements or additives, especially where the food is
a milk substitute. The method can be used to sort cells based on their
ability to bind to a Tir independent cell binding domain of an intimin
polypeptide. Polypeptides having Tir-independent intimin binding
activity can be used to produce a vaccine against a bacterial disease.
The present sequence represents a specifically claimed intimin C-terminal
Tir binding domain amino acid sequence, for use in the method of the
present invention.

Sequence 107 AA;
Query Match      18.9%; Score 530; DB 21; Length 107;
Best Local Similarity 99.1%; Pred. No. 8.3e-34;
Matches 106; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 255 QALALTPEPDPTTTDPDAANAASATKDQLTQEAQKPNQKVNIDANGNAIPSGELX 314
DB 1 QALALTPEPDPTTTDPDAANAASATKDQLTQEAQKPNQKVNIDANGNAIPSGELX 60

QY 315 DDIVEQIAQAQAGEVARQAVESNAQAQORYEDQHARRQELQLS 361
DB 61 DDIVEQIAQAQAGEVARQAVESNAQAQORYEDQHARRQELQLS 107

RESULT 4
AAY06213
XX AAY06213 standard; Peptide; 30 AA.
XX
AC AAY06213;
XX
DT 16-AUG-1999 (first entry)
XX
DE EPEC E. coli translocated intimin receptor N-terminal peptide.
XX
KW Tir; translocated intimin receptor; Hp90; enteropathogenic;
KW EPEC; infection; diagnosis; vaccine.
XX
OS Escherichia coli.
XX
PN WO9924576-A1.
XX
PD 20-MAY-1999.
XX
PF 10-NOV-1998; 98WO-CA01042.
XX
PR 12-NOV-1997; 97US-0065130.
XX
PA (UYBR-) UNIV BRITISH COLUMBIA.
XX
PI Devinney R, Finlay BB, Kenny B, Stein M;
XX
DR WPI; 1999-337712/28.
XX
PT New translocated intimin receptor useful for treating infection by
```

enteropathogenic or enterohemorrhagic Escherichia coli

Example 1; Page 37; 91pp; English.

The present sequence represents the N-terminal sequence of Tir (see also AAY06220), a novel translocated intimin receptor from an enteropathogenic Escherichia coli (EPEC) strain. The 78 kDa EPEC protein is secreted by the bacterial pathogen. Diagnosis of disease caused by pathogenic E. coli can be performed by use of antibodies that bind to Tir to detect the protein, or the use of nucleic acid probes for detection of nucleic acids encoding Tir. A kit for the detection of Tir-producing E. coli is provided. Also provided are a method of immunising a host with Tir to induce a protective immune response, and a method for screening for compounds which interfere with the binding of bacterial pathogens to their receptors.

Sequence 30 AA;

Query Match 5.9%; Score 165; DB 20; Length 30;
Best Local Similarity 100.0%; Pred. No. 5.3e-06; Mismatches 0; Indels 0; Gaps 0;
Matches 30; Conservative 0;

QY 2 PIGNLGNVGNHLLPPPLPSQTDGAAR 31
DB 1 PIGNLGNVGNHLLPPPLPSQTDGAAR 30

RESULT 5

ABB69655
ID ABB69655 standard; Protein; 934 AA.

AC ABB69655;

XX 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 35757.

XX Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

OS Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL13758.

XX New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -

XX Disclosure; SEQ ID NO 35757; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 934 AA;

Query Match 5.4%; Score 151; DB 22; Length 934;
Best Local Similarity 19.6%; Pred. No. 0.0074;
Matches 111; Conservative 74; Mismatches 218; Indels 164; Gaps 21;

QY 18 PAPPLPQTDGARGGTGHLISSTGALGSRSLSPLENSMADSVDSRDPLGPTNPRLA 77

DB 322 PAEDTTANRNGSAGSGSS---SKSGVIPKMR--KSLRHSQNLDD-----TEAAEVE 368

QY 78 AATSETCLLGFEVLHDKPLDILINTQIGPSAFREVEQADGTH----- 120

DB 369 SAAAEET---GGEQOSEHGNLDVIDQLPTPEAGNQSEASGIHNQSQLEADRPVQDED 424

QY 121 -----AAIGKNGLEVSVTLSPQWSSLSQIDTE----- 149

DB 425 DEEDDEEEEEEEVEGVGYIVNSAGSYEDTQIVAEDEDEITTEEEVDDDDDEIEBED 484

QY 150 -GKRFV--FTGGRGG-----SGHPWTVVASDIAEAARTRILAKLDPDNH 190

DB 485 LSESEFAQLIGELGGERQOPQRATNNAKTKNTTPAVAAASTTKATTATTS----- 537

QY 191 GGROPKVDVTRSVGVSGSIDDCVWSETHSTTNSVRSRDPKFWVSVGAIAAGLAGLAA 250

DB 538 SGYQQR---AAPHGQGAAPGAGGGGPRATRSSNSNN-----VVDYSIMPHLTVTAT 588

QY 251 TGIQAALATPEPDPPTTTDDQANAASATKQDLTQEAFFNPENQKVNIDANGNAIIPS 310

DB 589 T---TPHALTPQHQHQORQPPQPPSHQOQQVHQOQQPPQLL--PTHQFAHL--SSFVTPT 641

QY 311 GELXDDIVEQIAQQAQKEAGEVARQAVESNAQAQRYEDQARQOELQLSSG--IGYGL 368

DB 642 A-----AAAAAAAHYPPASAAAYFAQOQHQQHQOHHQOHHQOQAQOPHYH 692

QY 369 SSALIV-----AGGIGAGVTTALHRRNQPAEQT-----TTTTTIVVQQQTGGIPQHK 416

DB 693 SSVAVLHQPPPHFTSTGAGPPPALFQQAQAPQLTRYTPATATTAATFVPPQ-----QQQ 747

QY 417 VALMPQERRRFRDRDQSGSVASTHWSDSSEVSNVPAEVGAGNLSAHOPEHIYDEV 476

DB 748 VVYNPQQQHSLSRRSRG-----KTGSCVSSAAAAQQQHHQOQ-- 787

QY 477 AADPGYSVIONFGSGGPVTGELIGTPG 503

DB 788 -----QHSSAAAAVQQQLPPPG 805

RESULT 6

AAW68201

ID AAW68201 standard; Protein; 831 AA.

XX AAW68201;

XX 07-OCT-1998 (first entry)

XX M. catarrhalis strain O35E UspA1 antigen.

XX Moraxella catarrhalis; UspA1; UspA2; antigen; genetic vaccination;

XX vaccine; otitis media; sinusitis; lower respiratory tract infection;

XX immunity enhancer; immunoassay reagent.

XX Moraxella catarrhalis.

XX WO9828333-A2.

XX 02-JUL-1998.

XX 19-DEC-1997; 97WO-US23930.

XX XX

QY 179 TRILAKLPDNDHGGKQPKVDTRSVGVSGASGIDDDGVVSETHSTSTNSSVRSRDPKFWVSV 238
 DB 367 PNKICKTE---HVSASP---ASNSSASNNAGV-SGVSTSTQAIWQIVVARDKDTKWM 420
 QY 239 GAIAAGLAGLAATGIAQA-LALTPEDDPT-----TTDPQAAANAESA 281
 DB 421 TSLGMMNG-GLLGVPMGFLDTPPEPPAPSATPVTVEHVDLSCNPSPTDRDLSNTEPL 479
 QY 282 TKD-----QLTQEAEPENQKNDANGNAIPSGELXDDIVEQIAQAQAGEVAR 333
 DB 480 PIDNHLAQOIHRLDQSPMHSISHHTGDESN-----LVQHIKSEVIEAKHLAA 529
 QY 334 QQAVESNAQAQRYEDQHARQEEELQSSGIGYGLSSALIIVAGGIGAGVTTALHRRNQA 393
 DB 530 QHIALSQQAQHAHQHQHQHQHQ-----HQQQQQH 564
 QY 394 EOTTTTTTHVVQQTGGIPQHKVAMPQERRRFSRRDSQGSVASTHWSSSSEVNNPY 453
 DB 565 LHAQQLLAGSQLQQ-----QQQQQQQQHQHQHQQQQAAAAAAGVHGQHGHTV---H 616
 QY 454 REVGAR-----NLSAHOPEHHYDEVA-----A 478
 DB 617 ADIGGATVMEIDPSQIKHEPGMIITPEIVNMSSGHMDMYSNDSSESMIANGSPHPQK 676
 QY 479 DPGYSVIONFSGSPVTRGLICTPGGIGQSTYALLANGGLRLGMLTSGGETAVSSVN 538
 DB 677 EPHYNLDQHQHGLG---GSVCG-PQPG-----CAGGGGMSGAGSGSGEKDALKQN 725
 QY 539 AAPT 542
 DB 726 MSLT 729

RESULT 11

AAAG91741
 ID AAG91741 standard; Protein; 1488 AA.

AC AAG91741;

XX 26-SEP-2001 (first entry)

DE C glutamicum protein fragment SEQ ID NO: 5495.

XX Corynebacterium; amino acid synthesis; vitamin; saccharide;
 XX organic acid synthesis.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX 20-JUN-2001.

XX 18-DEC-2000; 2000EP-0127688.

XX 16-DEC-1999; 99JP-0377484.

XX 07-APR-2000; 2000JP-0159162.

XX 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX DR N-PSDB; AAH66960.

XX Novel polynucleotides derived from Corynebacterium bacteria, for identifying
 PT mutation point of a gene, measuring expression of a gene, analysing
 PT expression profile or pattern of a gene and identifying homologous gene

XX Claim 17; SEQ ID NO: 5495; 246pp + Sequence Listing; English.

XX

CC The present invention provides a number of nucleotide and protein
 CC sequences from the Corynebacterium glutamicum. These
 CC are useful for identifying the mutation point of a gene derived from a
 CC mutant of corynebacterium, measuring expression amount and
 CC analysing the expression profile or expression pattern of a gene derived
 CC from Corynebacterium, and identifying a homologue of a gene derived
 CC from Corynebacterium. Corynebacterium bacteria are useful for producing
 CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
 CC particularly L-lysine. The present sequence is a protein described
 CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX SQ Sequence 1488 AA;

Query Match 5.1%; Score 141.5; DB 22; Length 1488;

Best Local Similarity 18.9%; Pred. No. 0.078;
 Matches 106; Conservative 75; Mismatches 225; Indels 155; Gaps 19;

QY 48 SLF-SPLRNSMADSVDSRDIPGLPTNPRLAAATSETCLLGGFEVLHDKGPLDILNTQIG 106
 DB 778 ALFQPLADSLSSNESAEELVAGAGMGLVGNALATSGNWAAGWV----- 822
 QY 107 PSAPRVEVDAGTHAAIGEKNGLEVSVTLSPQEWSSLOS-ID---TEGKRFVPT----- 157
 DB 823 -----TVASSMGVLIIFTIIMKVRGSLVSGVDEAVTSVINRFLDTQVSSA 867
 QY 158 GGRGSGHPVTVASDIAEARTILAKLPDNDHGRQPKVDTRSVGVSGASGIDDDGVSV 217
 DB 868 GATSGDGMRRRAATGLGIGATHMVLNRDGGG-----SDSGSGSGSGSGSLGGEKAAG 923
 QY 218 EHTSTTNSVRSRDPKFWSVGAIAGLAGLAATGIAQALALTEPDDPTTTDPDQANA 277
 DB 924 LAKVTT-----VAGAGLVGYATD-----ALDNYADGVINGDGGAFAA 963
 QY 278 AESATKQDLTOEAFKNPENQKVNIDANGNAIPSGELXDDIVEQIAQAQKEA----- 328
 DB 964 GGDATVDG-----DYVADGDATASADANADFVDGVVDGAGRASFNSAAYSS 1009
 QY 329 -GEVARQAVESNAQAQRYEDQHARQEEELQSSGIGYGLSSALIIVAGGI-GAGVTTAL 386
 DB 1010 DGTTLDEGASVDAQGNPLHADGTFMSAAEAEMKMA-GLSSSGTWMEKSGVKSGIITTA 1068
 QY 387 HRRNQAPQTTTTTTHVQQQTGGIPQHKVAMPQERRRFSRRDSQGSVAST----- 440
 DB 1069 DVMDQSLASSVTESGLSKIPDTYG-----ADVSGAAGTVGTTGADYSA 1112
 QY 441 -----HMSDSSSEVNNPYAEVGGARNLSAHOPEHHYDEVAADPGYSVIONFSGS-- 491
 DB 1113 TDSSAGLNMSEALQSGTFMGALAGGSVS-SSDQAMMDAALQIAASQGLAPAGSIAGMEQ 1171
 QY 492 -----GPTVGRLLIGTP-----GQGIOTYALLANSGLRLGMLG 526
 DB 1172 LSAQATEAPAGKAGKQLGDLGSLNTQLASMGQGVGDSVNSAYA-----AG-----GMGGV 1223
 QY 527 TSGGETAVSSVNAAPTGPVR 547
 DB 1224 DVAGKVTAAQHLSQLVFCQIQ 1244

RESULT 12

AAW10660

ID AAW10660 standard; Protein; 580 AA.

XX AAW10660;

AC 20-JUL-1997 (first entry)

XX Endoglycoceramidase activator II.

XX Endoglycoceramidase; activator; glycolipid; enzyme.

```

XX OS Rhodococcus spp. M-777.
XX PN EP759470-A2.
XX PD 26-FEB-1997.
XX PF 28-JUN-1996; 96EP-0110513.
XX PR 29-JUN-1995; 95JP-0188466.
XX PA (TAKI ) TAKARA SHUZO CO LTD.
XX PI Ito M, Izu H, Izumi Y, Kato I, Kurome Y, Sano M;
XX WI; 1997-147519/14.
XX DR N-PSDB; AAT61050.
XX DN DNA encoding endo:glyco:cer:amidase activator polypeptide - for
PT prodn. of recombinant polypeptide, useful in glycolipid analysis
XX PS Claim 1; Page 13-15; 30pp; English.
XX CC DNA encoding endoglycoceramidase activator polypeptide can be used
CC for the prodn. of the recombinant polypeptide, which can be used
CC to research the intracellular function of glycolipids, because
CC it shifts the optimum pH for endoglycoceramidase II towards
CC neutral, allowing the latter to be used to hydrolyse glycolipids
CC even at pH 7.5.
XX SQ Sequence 580 AA;
Query Match 5.0%; Score 141; DB 18; Length 580;
Best Local Similarity 23.9%; Pred. No. 0.023;
Matches 84; Conservative 45; Mismatches 126; Indels 96; Gaps 16;
QY 4 GNLGNNVGNHLLPP-----APPLPSQTDGAARGGTGHLISSTGALGSRSLFSP 53
DB 215 GQFNDANFLFLPRASAPIVGTWAPTQCSPRDTAA-----GPLNAGAGPLATI 267
QY 54 RNSMADSVSDRIPLGPTNPSR--LAAATSETCLLGGFEVLHD-----KGPLDILN 107
DB 268 RQAVA-TVSYLDGPSAVTNGGFTLNATVPTDPSGQVQFTRDGDVGAPVDLVN---GK 323
QY 108 SAFRVEVQADGTHAAIGKNGLEV-----SVTLSPQEWSSLSQIDTEGKRRVF 159
DB 324 ASLTQSLTDGDYAEAKFLGAEFPNPSAAKTVTTSQDITTSVT----- 371
QY 160 RGGGHPMVTVASDIAEARTRILAKLPDNDHGGKQPKDVTDRSVGVGSASGIDDGWV 219
DB 372 --GPDH-----DAVRDQPVNLTAKVEPGVSGTVAFEVD--GTPVGTADVMDGAA 421
QY 220 HTSTNNSVRSDPKFVSVGAIAAGLAGLATGIAQALAL-----TPPDPTTDDP 271
DB 422 HTFTTNGTHR-----VIARYSG--AEGISPSVSLQPVSVTEAPADVATTIV 468
QY 272 DOANAASATKDLTQBAFKNPENOK-----VNIDANGNA 307
DB 469 DPIASTAGS-----PVTILRLDPADARGTVQFKLGDVLLGGPVRVDANGVA 516
RESULT 13
ABB59423
ID ABB59423 standard; Protein; 1472 AA.
XX AC ABB59423;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 5061.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

```

```

XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WI; 2001-656860/75.
XX DR N-PSDB; ABL03526.
XX DN New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
XX PS Disclosure; SEQ ID NO 5061; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABLI6176-ABLI30511), expressed DNA
CC sequences (ABLI01840-ABLI6175) and the encoded proteins
CC (ABB57737-ABB72072).
XX CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1472 AA;
Query Match 5.0%; Score 140.5; DB 22; Length 1472;
Best Local Similarity 21.0%; Pred. No. 0.092;
Matches 121; Conservative 75; Mismatches 240; Indels 139; Gaps 23;
QY 8 NNVGNHLLPAPPLPSQTDGAARGGTGHLISSTGALGSRSLFSLRNSMADSVSDR 67
DB 343 NVVSGTASVTAVPAAGCTEISAPGLEHQLVQGS-----SHLESSSG-----P 388
QY 68 GLPTNPSLAAATSETCLLGGFEVLHDKGPLDILNLTQIGPSAFRVEVQADGTHAIGEK 127
DB 389 AAVTPPATLSGSATTPLLQYSAAVSNPPPLQSQGTQSGAGT-----GASAAAGGA 440
QY 128 GLEVS--VTLSPQEWSS-----LQSIDTEGKRFVFTGGGSGHPMV 168
DB 441 GSTSFSSVSPDFTFSSAAATLVHQAQKQQLQQTTPK-----PSA 486
QY 169 TVASDIAEARTRILAKLPDNDHGGKQPKDVTDRSVGVGSASGIDDGWVSETHSTN 228
DB 487 TLSVEQSYFNSLASQ-----GVSPGVVQSPAGYA---QNPVAYSQTS--TSVCV 535
QY 229 RSDPKFVSVGAIAAGLAGLAATG-----IAQALALTPEPDPTTTTDPDANAABES 280
DB 536 SQYPNTYANV--FASGTAAGAGTAEQSQQPPQIRARVKLPPP----- 576
QY 281 ATKQLTOEAFKNPENQKVN-----IDANGNAIPSGELXDDIVEQIAQQAKEAGEVARQA 336
DB 577 ---SKIPASAVEMFGDNALNNIGVLDVQFGALDFG--TDDGFEPLPEKVGSGFSDGQ-- 629
QY 337 VESNAQAQRYEDQHARQEELQLSSGIGYGLSSALIVAGGIGAGVTTALHRRNPAREOT 396
DB 630 -----QQQQQDDYQSKSQQQQQVTV--LAAGLQSSQISDALNAGYTS---RSTQQQ 679
QY 397 TTTTHTTVVQQ-----QTGG-----IPQHKVLMFQRRRFRDRDSQGSVASTHSDS 445

```

Db 680 VSSAVNATIDQLTKSDPYGQTGGSGNAYQNAVQSSGASKTASGPTTAPGYSSTYANV 739
 QY 446 SSEVNPVAVGAGNSLSAHOPEHIIYDEVAADCHYVIONFSGSGPVTCRLITGTCQG 505
 Db 740 OSSVANSYQQQG-----YGSYQPS--YQQQAGSGAQ-----SGTGAVSGG-GGTATQN 785
 QY 506 IOSTVALLANSGLRLGMGLTSGGTAVSSVNA 540
 Db 786 IPVGGSSONSTSGNASSAYLTSGYSTPQSAVQSS 820

RESULT 14
 ABP43746
 ID ABP43746 standard; Protein; 985 AA.
 XX
 AC ABP43746;
 XX
 DT 26-FEB-2003 (first entry)
 XX
 DE 2 PUMH2 protein.
 XX
 KW Neuroprotective; immunomodulator; cancer; chromosome 2;
 KW cytostatic; anti-inflammatory; gene therapy; nutritional supplement;
 KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;
 KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;
 KW vulnery.
 XX
 OS Homo sapiens.
 XX
 PN WO200231111-A2.
 XX
 PD 18-APR-2002.
 XX
 PF 11-OCT-2001; 2001WO-US27760.
 XX
 PR 12-OCT-2000; 2000US-0687527.
 XX
 XX (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;
 PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;
 XX
 WIPI; 2002-426278/45.
 DR N-PSDB; ABQ60990.
 XX
 PT New polypeptides and their encoded proteins, useful as nutritional
 PT sources or supplements, or in gene therapy, particularly for treating
 PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or
 PT inflammation
 XX
 PS Claim 20; SEQ ID # 649; 357pp + sequence listing; English.
 XX
 CC The invention relates to 446 newly isolated polynucleotide sequences.
 CC The activity of polynucleotides of the invention may be described as,
 CC vulnery, neuroprotective, immunomodulator, cytostatic and
 CC anti-inflammatory. Compositions comprising nucleic acids of the invention
 CC are useful for treating a mammalian subject, or as nutritional sources or
 CC supplements. These are useful in gene therapy, particularly for treating
 CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,
 CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or
 CC inflammation. The nucleic acids and polypeptides are also useful in
 CC diagnostic and research methods. The sequences given in records
 CC ASP43544-ABP43989 represent polypeptides encoded by polynucleotides of
 CC the invention.
 CC NOTE: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 985 AA;

Query Match 5.0%; Score 139.5; DB 23; Length 985;
 Best Local Similarity 22.4%; Pred. No. 0.063;
 Matches 132; Conservative 65; Mismatches 221; Indels 171; Gaps 30;

QY 39 SSTCALGSRSLFPLRNMSADSVSRDIPGLPTNPSRLAAATSETCLGGEFVHLHDKGPL 98
 Db 85 SESEGLGVSMVEYVLSSSPADKLSRFRKG-----NFGTRDAET---DGPEKGDQKKG- 134
 QY 99 DIUNTOIGPSAFRVE-----VQADGTHAAI---GEKNGLEVSV-----TILSPQEWSSLOS 145
 Db 135 -----ASPFEDQNRDLKQGGDDDDSKINGRLPNMGMDADCKDFNRTPGSKQASPTEV 186
 QY 146 IDTEGKNRFVFTGGRGSG-----HPMVTVASD-----IAEARTRI-LAKLDPPDNHGG 192
 Db 187 VERLGNP-----TNPSEGLGPLNPMTANKPLVEFSNPETQNLDMAMEQVGLSLOFYPCN 242
 QY 193 RQPKVDVTRSVGVGSASGIDDDGVVSETHSTTNSVSRDPKFWVSVGAIAGLAGLAATG 252
 Db 243 QVPMDSGATVGLFDYNS-QQQLFORTNALTVQQLTAAQQQYALAAAQQPHIAGVFSAG 301
 QY 253 IAQALALTPEP-----DDPTTDDQANAASATKQDLTQEAFFKNPENQKVNIDANGNAI 308
 Db 302 LAPA-AFVPNPYIISAAPPDGTPTAAGLAAAT---LAGFAVVPPQY-----GV 348
 QY 309 PSGELXDDIVEQIAQAKAEAGEVARQAVESNAQAQRY-----EDQHARRQ 355
 Db 349 PWGVYVPAQLFQQ-QAAAAANNVTASQAASQAQPGQQQVLRAGAGORPLTPNQSQGQQA 406
 QY 356 EELQLSS-----GIGYGLSSAL-----IVAG-----GIGAGV-----382
 Db 407 ESLAAAAAANPTLAFQGGGLATGMPGYQVLAPTAYYDQTGALVVGFGARTGLGAFVRLMAP 466
 QY 383 TTAL--HRRNPAREQTTTTTHTTVVQQTGGI-----POHKVALMPQERRRFSRRD 432
 Db 467 TPVLISAAAAQAAAAAAGGTASSTLGTSTGLFRPIGTQPPQ-----QQQQPSTNLQ 519
 QY 433 SQGSVASTHWSDDSSSEVVNPVAVGVGAGNLSAHOPEHIIYDEVAADPGVSVIONFSGSG 492
 Db 520 SNSFYGSSSLTNSQS-----SSLFSHGP-----GQPG-STSLGF-GSG 556
 QY 493 PVTGRLITGTPQGGTOSTYALLANSGLRLGMGLTSGGTAVSSVNAAP 541
 Db 557 NSLGAAGISALSFGSS-----GGLTNGSGRYIS---AAP 588

RESULT 15
 AAB23860
 ID AAB23860 standard; Protein; 2411 AA.
 XX
 AC AAB23860;
 XX
 DT 17-JAN-2001 (first entry)
 XX
 DE Haemophilus influenzae adhesin (Hia) protein from type c strain API.
 KW Hia; adhesin; Haemophilus influenzae adhesin; NTHi; infection; vaccine;
 KW non-typeable Haemophilus influenzae; antiinflammatory; auditory;
 KW antibacterial; meningitis; epiglottitis; septicaemia; otitis media;
 KW diagnosis; immunogenic; antigen.
 XX
 OS Haemophilus influenzae.
 XX
 WO2000055191-A2.
 PD 21-SEP-2000.
 XX
 PF 16-MAR-2000; 2000WO-CA00289.
 XX
 PR 16-MAR-1999; 99US-0268347.
 XX
 PA (CONN-) CONNAUGHT LAB LTD.
 XX
 PI Loosmore SM, Yang Y, Klein MH;
 XX
 DR WPI; 2000-618897/59.
 DR N-PSDB; AAA92499.

THIS PAGE BLANK (uspto)